Applications of grazing-angle reflection absorption Fourier transform infrared spectroscopy to the analysis of surface contamination

A thesis submitted in the partial fulfilment of the requirements for the degree of

Doctor of Philosophy in Chemistry

at the

University of Canterbury
Christchurch
New Zealand

Michelle Hamilton

January 2007

Abstract

Cleaning validation of pharmaceutical manufacturing equipment is required by legislation. Generally, wet chemical techniques are employed using swabbing and/or rinse sampling methods. These are generally either selective and time consuming, or less selective and give results in a shorter period. The infrared reflection absorption spectroscopy (IRRAS) technique explored here attempts to deliver accurate, selective surface contamination information in real time to complement current methods and reduce down-time.

The IRRAS instrument used in this research is a Fourier transform infrared (FTIR) spectrometer coupled by an IR fibre-optic cable to a grazing-angle sampling head with a fixed incidence angle of 80°. The introduced flexibility permits collection of *in situ* spectra from contaminated surfaces. Calibration models are developed using the multivariate, linear partial least squares (PLS) statistical method.

The research focuses on sodium dodecyl sulfate (SDS), a model cleaning agent, on metal (aluminium and stainless steel) and dielectric (glass, EPDM and silicone) surfaces. The effects of surface finish are investigated for SDS on stainless steel. Calibrations for SDS and paracetamol in the presence of each other on glass surfaces are examined, as well as a common industrial cleaner (P3 cosa® PUR80) on polished stainless steel.

For the calibration sets in this thesis, RMSECV values were < 0.41 µg cm^{-2}, corresponding to conservative surface residues detection limits of better than ~0.86 µg cm^{-2}. However, RMSECV values depend on the calibration loading range, and the detection limits were typically ~0.2 µg cm^{-2} for loading ranges 0–2.5 µg cm^{-2}. These are below visual detection limits, generally taken to be 1–4 µg cm^{-2}, depending on the analyte and substrate. This shows that IRRAS is a viable method for the real-time detection and quantification of surface contamination by surfactants and active pharmaceutical ingredients on metals and dielectrics.

Acknowledgements

First and foremost I would like to thank my Ph.D. supervisors: Peter Harland and Bryce Williamson, since without your encouragement, guidance and friendship the completion of this thesis would not have been possible.

I would also like to thank the present and past members of the Harland-Williamson group: Ben, Josh and James, for your friendship and stimulating discussions during the course of my research. I have greatly enjoyed my time at the University of Canterbury this is in no small part due to the people I have met here; your friendships have been very much appreciated.

The Department of Chemistry is fortunate to have many talented technical staff, from whom I have been able to learn during my research. Much of the construction and repairs completed during this were carried out by the mechanical workshop boys; special thanks goes to Danny Leonard. The NMR section of my research would also not have been possible if it was not for the guidance of Martin Lee.

Thanks also go to Mary Thomson and Peter Melling for their support, especially in the initial stages of the project. Acknowledgement also goes to their company, Remspec Corporation, Massachusetts, for the provision of a graduate student scholarship and for the provision of equipment and software, without which this project would not have started. I would also like to thank the University of Canterbury Evans' Fund, the New Zealand Institute of Chemistry (NZIC) and the New Zealand Federation of Graduate Women (NZEGW) for the provision of funds which have allowed me to travel to a number of conferences both in New Zealand and overseas.

A final big thank you goes to my family and James, the task of completing this thesis would not have been possible if it was not for the love, support and patience you have shown me over the years.

List of publications

1. Grazing-Angle Fiber-Optic Fourier-Transform Infrared Reflection-Absorption Spectroscopy for the in situ Detection and Quantification of Two Active Pharmaceutical Ingredients on Glass. Perston, Benjamin B.; Hamilton, Michelle L.; Harland, Peter W.; Williamson, Bryce E., Thomson, Mary A.; Melling, Peter J. *Analytical Chemistry*, **2007**, *79(3)*, 1231-1236.

2. Fiber-Optic Infrared Reflection-Absorption Spectroscopy for Trace Analysis on Surfaces of Varying Roughness: Sodium Dodecyl Sulfate on Stainless Steel. Hamilton, Michelle L.; Perston, Benjamin B.; Harland, Peter W.; Williamson, Bryce E., Thomson, Mary A.; Melling, Peter J. *Applied Spectroscopy*, **2006**, *60(5)*, 516-520.

3. Grazing-Angle Fiber-Optic Infrared Reflection-Absorption Spectroscopy for in Situ Cleaning Validation. Hamilton, Michelle L.; Perston, Benjamin B.; Harland, Peter W.; Williamson, Bryce E., Thomson, Mary A.; Melling, Peter J. *Organic Process Research & Development*, **2005**, *9(3)*, 337-343.

4. Synthesis, structures and rac/meso isomerization behaviour of bisplanar chiral bis(phosphino-η5-indenyl)iron(II) complexes. Curnow, Owen J.; Fern, Glen M.; Hamilton, Michelle L.; Jenkins, Elizabeth M. *Journal of Organometallic Chemistry*, **2004**, *689(11)*, 1897-1910.

5. Mechanistic Studies on a Facile Ring-Flipping Process in Planar Chiral Ferrocenes under Ambient and High Pressure and Its Relevance to Asymmetric Catalysis. Curnow, Owen J.; Fern, Glen M.; Hamilton, Michelle L.; Zahl, Achim; van Eldik, Rudi. *Organometallics*, **2004**, *23(4)*, 906-912.

6. Absolute electron impact ionization cross-sections for the C_1 to C_4 alcohols. Hudson, James E.; Hamilton, Michelle L.; Vallance, Claire; Harland, Peter W. *Physical Chemistry Chemical Physics*, **2003**, *5(15)*, 3162-3168.

7. Oxide, Sulfide, Selenide, and Borane Derivatives of Indenylphosphines. Adams, Julian J.; Berry, David E.; Curnow, Owen J.; Fern, Glen M.; Hamilton, Michelle L.; Kitto, Heather J.; Pipal, J. Robert. *Australian Journal of Chemistry*, **2003**, *56(11)*, 1153-1160.

Contents

CONTENTS

List of abbreviations

ADC	Analogue to digital converter
API	Active pharmaceutical ingredient
CI	Confidence interval
CLS	Classical least squares
DL	Detection limit
DOF	Degrees of freedom
DTGS	Deuterated triglycine sulfate
EPDM	Ethylene propylene diene monomer
FDA	United States Food and Drug Association
FTIR	Fourier transform infrared
GC-MS	Gas chromatography – mass spectrometry
HPLC	High performance liquid chromatography
IE	Imbedded error
ILS	Inverse least squares
IR	Infrared
IRRA	Infrared reflectance absorbance
IRRAS	Infrared reflection absorption spectroscopy
MCT	Mercury cadmium telluride
MSE	Mean-squared error
MSEC	Mean-squared error of calibration
MSEP	Mean-squared error of prediction
NMR	Nuclear magnetic resonance
PCA	Principal component analysis
PCR	Principal component regression
PEG	Polyethylene glycol
PLS	Partial least squares
ppb	Parts per billion
PRESS	Predicted error sum-of-squares
RA	Reflectance absorbance
RE	Real error
RID	Reflective index detector
RL	Residue limit
RMSE	Root mean-square error
RMSEC	Root mean-square error of calibration
RMSECV	Root mean-square error of cross validation
RMSEP	Root mean-square error of prediction
SDS	Sodium dodecyl sulfate
SEM	Scanning electron microscope
SR	Spectral region
TOC	Total organic carbon
UV	Ultra-violet
XE	Extracted error

Introduction

The aim of the research presented in this thesis was to investigate whether infrared reflection absorption spectroscopy (IRRAS) could be used in the identification and quantification of trace surfactant and active pharmaceutical ingredients on dielectric and metal surfaces with potential applications to cleaning validation in the pharmaceutical industry. This research was motivated by REMSPEC Corporation, Massachusetts, which sees a need for a technique that is easy to use and can quickly quantify surface contamination in the pharmaceutical industry. The current issues in cleaning validation are related to time and cost. Most cleaning validation is carried out using wet chemical techniques, which are both time consuming and costly; there is also the requirement for sufficiently trained technicians to run the specialised equipment and to take samples. It is postulated that the IRRAS technique could be used to dramatically reduce both validation time and cost.

The first three chapters in this thesis are devoted to cleaning validation, infrared spectroscopy and chemometrics.

The first chapter discusses the origins of cleaning validation and its relationship to surface contamination. The regulations and guidelines that have been implemented in industry, particularly in the United States of America, are considered. There is a discussion on the analytical techniques that are currently used for cleaning validation and possible future directions for both cleaning validation and analytical techniques. Particular attention is paid to the application of FTIR spectroscopy in cleaning validation, since this is the analytical method employed in this research.

Following on from the first chapter, Chapter two is devoted to infrared spectroscopy. The first section of this chapter covers the basics of the FTIR technique, with a brief explanation on how the instrument works. Subsequent sections present detailed

considerations of the origins of the technique and its applications to both metallic and dielectric substrates. The theory behind the IRRAS technique is developed, with emphasis placed on the determination of the ideal experimental conditions to produce the best IRRA spectra.

The third chapter is devoted to chemometrics, which is the general name given to the process of extracting chemical information using mathematical or statistical methods. Due to the complex nature of the IRRA spectra collected during this research, traditional univariate quantitative analysis was not appropriate. To overcome issues associated with univariate analysis, multivariate analytical methods were investigated. The development of the multivariate methods – classical least squares and inverse least squares – are discussed. This methodology is then extended using factor analysis to more complex methods of principal component regression and partial least squares. Particular attention is given to the techniques used in this thesis.

Chapter four covers the instrument used in the collection of the spectra throughout this research. There is a description of the analytes and substrates that were used during this project. The experimental techniques used to prepare and quantify the samples are also discussed.

The remaining five chapters present and discuss the results. Quantification of SDS films on aluminium and stainless steel substrates are covered in Chapters five and six, respectively. Chapter six also includes an investigation into the effect of the surface roughness of stainless steel substrates. Chapter seven presents data for SDS on glass and combinations of SDS and paracetamol on the same substrate. Results for SDS on flexible (polymer) dielectric substrates, EPDM and silicone, are presented in Chapter eight. Chapter nine reports an investigation of an industrial multi-component cleaning agent, P3 cosa® PUR80, on polished stainless steel substrates.

The last chapter in this thesis is a summary of the general conclusions drawn from this work and possible future directions for further research.

Chapter 1

Cleanliness Validation

1.1 Cleaning validation

The prevention of contamination is important in any industry producing high-quality products, but is especially important in industries where the products are consumed, such as in drug and food processing and manufacturing facilities. Contamination in an industrial manufacturing situation could come from a number of sources, which can include: contaminated starting materials; poor cleaning of equipment; lack of training of operating staff; and non-compliance with standard operating procedures (SOPs). In many situations, even a slight degree of contamination of a product could result in a devastating cost to the manufacturing company through down-time in manufacturing operations while the source of the contamination is found and resolved. In many industries there is also the potential for legal action and consumer backlash, which can damage a company's reputation and future development. Therefore, the prevention of contamination through the use of control measures is a crucial issue for many manufacturing companies.

In order to help guide companies in this area, governments have empowered a number of organisations to regulate the industries concerned. Organisations such as the United States Food and Drug Association (FDA), have implemented regulations and testing protocols intended to preclude the manufacture of contaminated products.[*] One set of such regulations relates to cleaning validation, and was introduced in the United States in the early 1990s.

The Active Pharmaceutical Ingredients Committee (APIC) of the European Chemical Industry Council (CEFIC), an internationally recognized voice on Good Manufacturing

[*] Regulations and protocols referred to in this thesis relate to the situation in the United States of America unless otherwise stated.

Practice (GMP) and Regulatory Affairs, defines cleaning validation as, "the process of providing documented evidence that the cleaning methods employed within a facility consistently control potential carryover of product (including intermediates and impurities), cleaning agents and extraneous material into subsequent product to a level which is below predetermined levels."[1] The introduction of these requirements led to a large scale reassessment of the methods and tools used to carry out cleaning validation. Some important questions were raised, including: what *is* surface contamination?; when should cleaning validation be carried out?; what are the acceptable limits for contamination (how clean is clean)?; and how can contamination at the level of these limits be detected?

1.2 What is surface contamination?

Surface contamination can generally be classified into three categories. [2]

- Particulate contamination: this generally refers to particles such as dust, hair and fibres.
- Thin film or molecular contamination: this can cover some of, or the entire surface of, the exposed equipment and can be organic or inorganic. These types of contamination are of most concern for cleaning validation since their sources are likely to be from the manufacturing process itself, with examples including grease, oils and surfactant/chemical residues.
- Microbial contamination: this refers to any unwanted organism growing on the surface or on the residues left on the surface, and includes biological cultures, spores or bacteria.

The work reported in this thesis is primarily focused on thin film or molecular contamination and its relevance to the pharmaceutical industry. Particulate contamination is an issue, but should be easily removed in the cleaning process. Microbial contamination is of great importance, especially for the food industry. But the techniques used to determine microbial contamination are the same as those used for monitoring

sterile environments; they are well established and do not require special development.[3] The emphasis in the pharmaceutical industry is to prevent microbes getting on or into equipment in the first place.

1.3 When should cleaning validation be performed?

A large number of documents have been published that outline protocols that should be followed when developing a cleaning validation strategy for an entire plant or single product run.[1, 3] These documents indicate that, as soon as practicable after the end of a product run, the equipment must be cleaned according to SOPs, which should be documented by the company. Once the equipment has been cleaned, cleaning validation should occur as soon as possible afterwards, leaving the equipment ready for the next production run or for storage.

In many production plants, multiple products are manufactured using the same set of equipment. In order to prevent cross-contamination, the equipment needs to be cleaned and validated between each product run. Different levels of cleaning validation may be required; in general, the greater the potential for contamination in the final product the more stringent the validation protocols.[4] If, in the early steps of production, one intermediate product is combined with another, cleanliness may not need to be as stringently validated as would be the case in a situation where the final step requires a change of equipment.

The requirement for cleaning validation may also depend on subsequent processes. For example, if the primary manufacturing process is followed by purification, equipment cleaning validation may not need to be so rigorous. The nature of the compounds being used or synthesized in a production process also has a large impact on the cleaning validation requirements; if these compounds are particularly toxic or harmful there may be a need for more stringent cleaning and validation protocols.

In general, a company will want to minimise the amount of down-time and cost of cleaning validation, but this has to be balanced against maintaining an environment free from contamination and having the required amount of paper work to prove that cleaning is being carried out to a sufficient level.

1.4 Determination of residue limits

The determination of acceptable residue limits in drug manufacture is dependent on a number of factors, including the chemical and physical properties of the materials, where they are used in the production process, what production equipment they come into contact with, the product batch size and how the drug is administered (injected, ingested or in ointment form).

An individual target compound's physico-chemical properties, such as solubility, fluorescence, degradation products and rate of decomposition, are important in determining its residue limits. But there is a general consensus that residue limits should primarily be based on pharmacological activity or toxicity.[3, 5] It is reasonably easy to determine a compound's solubility and fluorescent properties, but characterising its physiological effects is more difficult. In particular, toxicity is one of the most difficult factors to determine accurately, since long-term toxicity data for humans are limited. Most acute and chronic toxicity studies have been performed on animals and the results have been extrapolated and adjusted for humans. Some indices that have been calculated are: no-observable-effect level (NOEL), lethal dose (LD50) and acceptable daily intake (ADI).[3] Each of these has its advantages and disadvantages. A NOEL is based on a two-year chronic toxicity study, where the highest dose with no toxic effects is determined. This method is limited because the values so determined are dependent on the manner and duration of exposure, the animal species tested, and the endpoint measure (what is the definition of an observable effect). LD50s are based on acute toxicity studies, which estimate the lethal dose for 50% of the tested animals (usually rats or mice) following a single dose of a compound. This has the advantage that a lengthy study is not required,

but it does not give any insight into long-term effects of a compound nor does it account for differences between the tested animals and humans. LD50 values are mostly used for non-active ingredients, such as cleaning agents and sanitizers. Finally, an ADI is the amount of toxicant, in mg kg^{-1} of body weight per day, that is not anticipated to result in adverse effects after chronic exposure of the general population of humans.[6] This is obviously difficult to determine, although, they can be estimated using LD50 values and applying a safety factor between 1×10^{-6} and 10^{-7} day^{-1}; these values are usually associated with ingested non-pharmaceutical products.[7]

Acceptable residue limits are also influenced by the stage at which the compound enters into, or is formed in, the sequence of steps involved in the production of the pharmaceutical. The closer the compound is towards the end of the production process, the higher the risk of contamination of the final product and the more stringent will be the requirement for lower residue limits. If there is a purification process in subsequent steps, then acceptable residue limits may be higher for preceding steps.

The risk of product contamination by surface compounds is related to the surface area of the vessels used in the manufacturing process. If two different products are made or processed in the same sets of equipment, the chances of cross contamination are higher than if there is only one piece of common equipment used for the manufacturing of each compound. Therefore, the residue limits for the contaminating compound in the situation where there is a large amount of common surface area between two subsequent production runs needs to be lower than for the situation where there is only a small amount of common surface area and the contamination can only come from a few sources. The production size and product dosage of the next compound to be processed in the equipment also influences the acceptable residue limits. Larger batch sizes and smaller dosages will minimise the possibility of product contamination and result in the acceptance of higher residue limits.

The above outlines do not represent a complete description of all factors that need to be considered when determining acceptable residue limits. Furthermore, it is clear that every

manufacturing plant is going to have unique problems due to having different processes involving different compounds. Because of this, the guidelines from organisations such as the FDA do not define the maximum residue limits for any compound on any piece of equipment in the production process. Instead, they state that residue limits should be established by the manufacturers since they have the expertise on the compounds, equipment, reagents and other information that may be relevant to determining the potential for contamination. The FDA *does* state that the residue limits set should be practical, achievable, justifiable and documented.[5]

Fourman and Mullen[8] have presented a case study on how residue limits can be determined in a processing plant that handles a number of compounds exhibiting different physical properties and utilising equipment with a range of product batch sizes. They presented three ways to determine residue limits (which they refer to as acceptance limits) based on:

(i) a requirement that no more than 10 ppm of any product will appear in any other product;
(ii) the pharmaceutical dose method; and
(iii) the policy that no residue is visible on a cleaned surface.

1.4.1 The 10 ppm method

The 10 ppm method has its origins in the US food industry and it is also mentioned in the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) quality guidelines.[9-15] The FDA has found that some US companies have not employed this guideline correctly and that the 10 ppm measurement needs to be implemented with care, the degree of which may be dependent on the stage of the manufacturing process.[1]

Equation (1.1) is the formula used by Fourman and Mullen[8] to calculate the residue limit (RL) for a given tested area of a compound A being followed by a batch of compound B, where *R* is equal to 10 mg of the active ingredient in compound A per kg of compound B.

M is the mass in kilograms per batch of the final mixture of compound B; *A* is the surface area of the equipment that is common to both manufacturing processes; *a* is the tested area, which, in the referenced article, was 4 in² (~ 26 cm²).

$$\text{RL} = R \times \frac{M}{A} \times a \qquad (1.1)$$

For example, if a batch of 15 kg of a product B is processed in equipment that shares 25,000 cm^2 with product A, then R = 10 mg would give an acceptable RL for product A in a tested area of 26 cm^2 of 156 μg;

$$\text{RL} = 10 \text{ mg kg}^{-1} \times \frac{15 \text{ kg}}{25{,}000 \text{ cm}^2} \times 26 \text{ cm}^2$$

$$= 156 \text{ μg} \qquad (1.2)$$

Note that the 10 ppm level is not necessarily internationally recognised; the Australian New Zealand Food Standards Code state the maximum acceptable level (ML) is, "the maximum amount of a specified contaminant, or specified natural toxicant, which is permitted to be present in a nominated food expressed, unless otherwise specified, in milligrams of the contaminant or the natural toxicant per kilogram of the food (mg kg^{-1})."[16] The MLs for a large number of compounds have been determined and are listed in the Australia New Zealand Food Code.[16]

1.4.2 The pharmaceutical dose method

In the pharmaceutical dose method, it is assumed that less than 0.001 of a typical dose of one component will appear in the maximum daily dose of any other pharmaceutical. This method assumes that pharmaceuticals at 0.1 of their typical dose are non-active. To arrive at the 0.001 value, safety and robustness factors of 0.1 each are also included.[8] There is some debate about the appropriate values to use with this method; some papers use the typical dose while others use the maximum therapeutic dose, which is more appropriate

to a worst-case scenario.[1, 8, 17] Other contaminants can also be treated in this way, with the acceptable residue limits being determined according to toxicity, for example by using the acceptable daily intake (ADI) instead of the normal therapeutic dose.[1]

Equation (1.3) describes how the residue limit (RL) can be calculated where f is the fraction of the daily dose (in mg day^{-1}) of the active ingredient in compound A, J is the maximum or typical number of doses of B taken per day and K is the number of doses per batch of compound B.[8]

$$\text{RL} = \frac{f}{J} \times \frac{K}{A} \times a \tag{1.3}$$

For example, consider a case where the contaminating product is a tablet that contains 50 mg of active ingredient A with the maximum intake of three tablets per day, and product B has a daily dose of six tablets per day. In a single batch of product B there are 30,000 tablets, the common surface area is 25,000 cm^2 and the tested area is again 26 cm^2. Therefore the acceptable RL in a tested area of 26 cm^2 would be 780 μg;

$$\text{RL} = \frac{0.001 \times 150 \text{ mg/day}}{6 \text{ tablets/day}} \times \frac{30000 \text{ tablets}}{25000 \text{ cm}^2} \times 26 \text{ cm}^2$$

$$= 780 \text{ μg} \tag{1.4}$$

1.4.3 The visible inspection method

The last of Fourman and Mullen's methods is visual inspection and is based on the precept that cleaned surfaces should appear to be clean. Studies have shown that most compounds are visible on a surface at ~ 4 μg cm^{-2},[8] although this reduces to ~1 μg cm^{-2} if a light is used to illuminate the surface.[18] For some compounds, the RLs calculated by the 10-ppm and pharmaceutical dose methods are larger than 4 μg cm^{-2}. It seems inappropriate that a surface can be declared clean if contamination is visible, and in such cases the requirement that the surface is visually clean is considered more acceptable.[8]

There has been some research into the use of visible residue limits as the only criterion for establishing whether a surface is clean or not.[19-21] There are a number of factors that need to be considered in this respect; they include the distance and angle of viewing, along with the position of the analysed surface and the amount of light falling on it. Sufficiently trained personnel are a requirement for this approach since validation would be completely dependent on their interpretation of what they see. In short, visual validation may be considered acceptable between batches of the same drug, but is unlikely to be acceptable for other cases. Regulatory bodies prefer analytical methods that are less susceptible to subjective interpretation and they are unlikely to accept a cleaning validation plan that is based solely on visual acceptance levels.

1.4.4 Residual solvents and cleaning agents

Many guidelines mention the need for residue limits for cleaning agents and solvent residues.[1, 3, 5, 12] Since cleaning agents are not integral to the manufacturing process, they should be selected to minimize contamination and to be easily removed; but the surface should still be checked for contamination from them. This can be difficult since cleaning agents are usually a mixture of compounds and their manufacturers are commonly unwilling to provide specific composition information. The difficulty then becomes one of determining the components for which it is necessary to establish an acceptable RL. There are number of approaches including: consideration of the most toxic component (e.g. smallest LD50); the component that is the predominant constituent of the formulation; whole-product analysis; and the most strongly adsorptive component, which is the one most difficult to rinse from a surface.[22, 23]

1.4.5 Other considerations

The determination of the overall acceptable RL for a compound can be carried out by calculating and comparing the RLs for each of the above methods. The lowest RL is usually accepted since it presents the best allowance for safety and error margins. But

there are still some limitations, the most obvious being that contamination may not be uniformly distributed. The common way to alleviate this problem is to use point testing to make sure that contamination in these areas is well within safety guidelines.[18]

Microbiological contamination is also subject to cleaning validation. The corresponding RLs are related to the sampling method (described in the next section) and are usually about 10 colony-forming units per 25 cm^2 for surface sampling or 100 colony-forming units per mL for rinse sampling.[3] As mentioned earlier, the methods for detecting microbiological contamination are generally well established but difficulties are still experienced since levels can change over time; this is especially important in the food industry.

1.5 How to sample surfaces for residues

Once acceptable RLs have been decided, appropriate methods need to be employed to determine the contamination levels actually present on the cleaned equipment. Such methods are usually categorised as direct or indirect, depending on whether the measurements are made directly or indirectly from the surface of interest. The FDA has outlined general methods of both types. Each has its advantages and disadvantages, but, in general terms, direct surface sampling is favoured since solubility and build-up issues are more easily controlled and monitored.[5] Other methods, such as coupon and placebo sampling (*vide infra*), are not commonly used and may need to be supported by direct surface or rinse sampling.

1.5.1 Swab sampling

Direct surface sampling can be carried out in a number of ways, but the most common and widely accepted is swabbing. This involves wiping a predetermined area of the equipment with a swab that has been moistened with a solvent determined by the contaminating compound. Usually the surface is wiped with one side of the swab using a

certain number of strokes, then the swap is flipped and the surface is wiped at 90° to the first series of stokes. This process can be extended, for example by repeating the process at diagonal angles. Generally, the swabbing proceeds from less contaminated to more highly contaminated areas in order to prevent recontamination by the material already collected on the swab.[17] The swab head is then cut from the handle and immersed in a set amount of "recovery solvent". The concentration of contaminant in the recovery solvent is then determined by analytical techniques and the amount of contamination on the swabbed surface is back calculated.

Amongst the advantages of the swabbing method is the fact that insoluble or slightly soluble residues on the surface are more readily removed by physical "rubbing" than is the case in (for example) rinsing. It also permits direct sampling from accessible, but hard-to-clean locations (although it can also be at a disadvantage for locations that are especially hard to clean or difficult to access). Generally, small sampling areas are used to determine local RLs, which are then extrapolated to estimate the level of contamination over the entire contact surface. This can lead to problems in large-scale reactors where the residues are not uniformly spread across all contact surfaces.

Issues that need to be considered when using the swabbing technique include the physical properties of the swabs, recovery levels and operator procedures. The swab material must not damage the surface or leave fibres behind; but most importantly it must not leach compounds that can interfere with the analytical procedures. One potentially significant interferant that can lead to problems of the last type is glue used to attach the swabbing head to the handle. To preclude this problem, thermal adhesion treatments are preferred over glues.

Recovery levels are determined by the solubility of the compound in the swabbing solvent, the wiping procedure and the physical nature of the surface. Ideally, they should be as close as possible to 100%, but greater than 70% is considered reasonable and as low as 50% is sometimes obtained. Lower values are generally considered unacceptable and to require improved procedures.[3, 24] To compensate for imperfect recovery, studies must

be carried out to determine appropriate factors to correct the calculations for the actual RL. There are two major reasons for imperfect recovery. Firstly, not all of the contamination on a surface may be collected by the swabbing process; secondly, not all of the contamination on the swab may be passed to the recovery solution. The first of these can be influenced by the type of surface and its roughness, as well as the type of swab and the solvent. A second swab can be used to improve recovery, with the results from both swabs being combined. The second swab can employ a different solvent, with appropriate consideration of residues and toxicity. A common secondary solvent used after water is ethanol, since it dissolves many pharmaceutical compounds, is reasonably non toxic and evaporates readily leaving no residue.

1.5.2 Rinse sampling

An indirect sampling method is rinse sampling. A small sample of the solution collected from the last rinse cycle of the cleaning process is analysed for the compound of interest and the RL is back-calculated according to the volume of solution and the contact area.

There are generally two assumptions inherent in this method. The first is that the target residue is efficiently extracted into the rinsing solution. The second is that all parts of the contaminated surfaces are cleaned equally. The main reservations expressed by the FDA and other regulatory bodies about rinse sampling relate to these assumptions. Since the surface residues are not measured directly, the analyst cannot be sure that unacceptably high levels of residues have not been left in some areas of the equipment. For example, a very poor solvent will result in low contamination of the final rinse even if large amounts of residues have been left on the surface. There is an additional risk of system failure, where an (otherwise) entirely adequate procedure is incorrectly applied. This could mean that a dirty reactor is declared clean and used for some time before the malfunction is detected.

Even with these concerns, rinse sampling does have a number of advantages when implemented correctly and with adequate safety measures. One fairly significant

advantage is the ease of collecting a part of the final rinse solution drained from the equipment. Another is that it allows evaluation of residues from all parts of the surface irrespective of the difficulty of reaching them with a swab. This makes rinse sampling ideal for clean-in-place (CIP) systems,[†] sealed systems or large-scale equipment that is difficult to disassemble.

1.5.3 Placebo sampling

The placebo sampling method involves manufacturing a batch of placebo (the product minus the active ingredient or contaminant of interest) in the cleaned equipment. The final placebo is then tested for the excluded component. There are several potential problems with this method. The first results when contamination occurs at one point in the process, so the placebo products that are formed first are likely to be more contaminated than those that follow later.[25] Secondly a lot of material is wasted in manufacturing the placebo. Thirdly, the levels of contamination in the placebo may be difficult to detect.

1.5.4 Coupon sampling

Coupon sampling methods involve the use of manufacturing equipment whose components can be removed and analysed for residues. This approach is useful when incorporated with analytical techniques that can make measurements directly from the surface. The main problem with the coupon sampling is that the removable components are usually flat and come from locations that are not hard to clean; they are therefore not representative of a worst case scenario.

[†] CIP systems are used in large scale equipment which is difficult to or cannot be dismantled and moved into a cleaning facility; instead they are cleaned where they are assembled. Cleaning is usually performed with a spray device.

1.5.5 Other considerations

One of the major disadvantages of the swabbing and other process (like rinse sampling) is that once the samples are collected they must be sent to a laboratory for analysis (see the next section). Until the results are returned from the laboratory, the equipment cannot be declared clean and should not be used. This may take some time – and time is related to profits.

Another important consideration is the establishment of an appropriate and effective protocol by which the cleaning is related to its validation. One such protocol is, "test until clean", where the cleaning and testing continue in an iterative cycle until the RL falls below the acceptable level. This has been deemed by the FDA to be generally unacceptable because it indicates a lack of control over the cleaning process.[5]

1.6 Surface residue detection methods

Once a sample has been collected, it must be analysed to determine the amounts of residues. Detection methods are (like sampling methods) categorised as indirect or direct. They are also either specific or non-specific. Specific methods, which permit identification and, possibly, quantification of a single species are preferred by regulatory authorities.[24] The ideal method is one that is rapid, free from interferences and distinguishes between compounds quantitatively, at very low levels with a high degree of confidence.

There are a large number of techniques that can be used for detection. The technique chosen for a particular component is dependent on the sampling method, the physico-chemical properties of the component, the RL and the required detection range. Some of the common specific techniques are: high performance liquid chromatography (HPLC); other chromatographic methods (such as liquid, gas and thin-layer chromatography coupled to detectors like mass or UV spectrometers); ion chromatography (especially for

potassium and sodium); atomic absorption spectroscopy; inductively coupled plasma (ICP) atomic emission; FTIR; and capillary electrophoresis. In addition, there are other less specific techniques, which include: total organic carbon (TOC); pH; acid-base titrations; conductivity; contact angle; spectroscopic methods using visible, IR or UV radiation to illuminate a sample surface; optically stimulated electron emission (OSEE); and optical particle counters. All of these methods have advantages and disadvantages that relate to cost, operator training, sensitivity and processing time. Two of the most commonly used techniques are considered below.[26] More complete descriptions of these and other techniques can be found in a number of reviews.[2, 3, 18, 24, 25, 27, 28]

1.6.1 High-performance liquid chromatography

The most commonly used analytical technique is HPLC, which can be combined with a variety of detectors depending on the compounds of interest. HPLC separates compounds according to their physio-chemical properties and the type of column used in the instrument. Since a compound with a stronger affinity for the solid phase of a column will take longer to pass through, the compounds in a sample will be separated. A detector placed at the end of the column detects the compounds as they exit the column. Analytes can be recognised and characterised by their retention time, which is independent of concentration. Depending on the detector used, the area under the retention peak can be used to determine concentration.

Detectors that can be coupled to HPLC systems include ones that measure fluorescence, ultraviolet (UV) absorbance, electrochemical potentials, refractive indices, conductivity and evaporative light scattering. Of these, UV detectors are the most common; most compounds have at least one UV-active chromophore and the method does not alter the sample, which can be important if other tests are to be performed. UV detectors can be very sensitive, are reasonably inexpensive, readily available, simple to use and relatively temperature independent. Additionally, there is no need for any pre- or post-column reactions and the analysis can be performed directly on the sample solution (although

some analytes may require pre-concentration). The UV spectra can also be analysed to give information about the nature of the detected compounds.

Unfortunately, not all compounds absorb UV radiation at readily accessible wavelengths. In the context of this thesis, the most significant examples are surfactants, which are used in the pharmaceutical industry as components in cleaning agents. To detect such compounds, a different type of detector is needed. One example is evaporative light scattering detectors (ELSDs). ELSDs are referred to as universal detectors since they respond to all non-volatile compounds leaving the HPLC column. They are simple, versatile and rugged; they should also be free of mobile-phase effects such as baseline drifts due to evaporation prior to detection (although this leads to complications since there are limited choices of suitable buffer salts).[3, 24, 25] In the literature there are also examples of surfactants being detected using gas chromatography coupled to a mass spectrometry (GC-MS), or by FTIR.[27, 29]

HPLC systems do have some drawbacks. Even though the detection methods are fast and reasonably inexpensive, the actual separation in the columns can be time consuming and an HPLC instrument, including its columns, is relatively expensive to both purchase and operate.

1.6.2 Total organic carbon analysis

Total organic carbon (TOC) analysis is a widely used technique, even though it is non-specific. It involves the oxidation of organic carbon to carbon dioxide, which is then detected by either IR spectroscopy or conductivity measurements.[23] TOC is rapid and very sensitive, permitting measurement of an analyte in the ppb range.[3] The cost is moderate, with the possibility of integration into on-line analysis, although this is only suggested for water quality analysis.[22] Since TOC is non-specific it is a good technique for detecting cleaning agent residues, which are likely to contain a number of compounds in possibly unknown concentrations, all or some of which may be difficult to measure using HPLC.

There are two distinct disadvantages of TOC. The first, that the analyte must contain carbon, is not usually an issue in the pharmaceutical industry. The second is that the compound needs to be water soluble, although solubility in the low ppb range is all that is required. Organic solvents cannot be used since their carbon content would confound the analysis.[3, 24, 25] This leads to one of the major concerns with the use of TOC; since all organic carbon is detected, independent of its source, the method is sensitive to incidental contamination, which can arise from sources such as the atmosphere, the swab itself and poor laboratory procedures.[3, 25] This can lead to an overestimation of surface contamination.

1.7 Fourier transform infrared spectroscopy and cleaning validation

Fourier transform infrared (FTIR) vibrational spectroscopy is a useful tool in the identification of molecules since it interrogates both the types of chemical bonding and the molecular structure. It can be applied to all phases and can be collected from any molecule that undergoes a change of electric dipole moment during vibration.

The ability to detect a wide range of compounds would appear to make FTIR a useful tool for cleaning validations, especially when combined with reference library data. But the ease with which active pharmaceutical ingredients (APIs) are detected using UV spectroscopy has meant that FTIR has traditionally been limited to the detection of cleaning agents, which are difficult to measure using HPLC and UV spectroscopy. Generally, FTIR has not been used as an independent method because of difficulties in measuring spectra *in situ*.[25] This can be alleviated by the use of coupon sampling (mentioned earlier) or measuring spectra obtained from swabbing and rinse solutions, although non-linear (non-Beer's law) type response of FTIR can lead to difficulties in producing a reliable calibration.[25]

Even with these difficulties, the potential of FTIR as a method for the analysis of surface residues has been supported.[23] Its implementation has been facilitated by the availability of commercial grazing-angle FTIR attachments and the development of mid-infrared fibre optics, which allow spectra to be taken directly from a surface. But some people in the industry consider that infrared fibre optics have not yet been developed enough to be used for cleaning validation.[24] Due to perceived limitations, FTIR is currently limited to screening and rinsibility studies, where grazing-angle FTIR[30] has been used to determine what compounds are selectively adsorbed to the surface and whether a reagent (usually a cleaning agent) has been completely washed from a surface. This has been particularly useful for single- or whole-component analyses.[31] In one rinsibility study, TOC was combined with grazing-angle FTIR to achieve a semi-quantitative calibration between 1.45 – 0.952 μg cm^{-2} using IR peak heights for an alkaline cleaner.[29]

Since coupon sampling methods are not encouraged by the FDA, the development of direct-sampling FTIR detection methods is desirable. The concept has developed of using FTIR with a grazing-angle sampling head and infrared fibre optic to collect spectra, *in situ*, directly from a surface. The spectra could then be compared with a calibration model in real time in order to determine the type and amount of contamination on the surface. The major advantage of such a technique would be that results are achieved quickly with minimal down-time between production runs.

A couple of prototype FTIR instruments are under development for *in situ* FTIR cleaning validation in the pharmaceutical industry; one by Sensiv Incorporated (Waltham, Massachusetts) and the other by Remspec Corporation (Sturbridge, Massachusetts). The Sensiv instrument, based on a non-contact reflectance probe coupled to an IR fibre optic cable, has been used to measure sodium lauryl sulfate (also known as sodium dodecyl sulfate or SDS) on stainless steel down to 5 μg cm^{-2}. The same researchers also reported measurements of Ibuprofen over the range 5–10 μg cm^{-2}, which corresponds to ~0.1% of typical therapeutic doses.[32] The Remspec system, which utilises an IR fibre optic to transmit radiation from a spectrometer to a grazing-angle sampling head, is the one that has been used in work reported in this thesis. Prior to this work, this instrument had

permitted measurement of fluorocarbon lubricant used on computer hard drives down to 0.2 μg cm^{-2}, and detection of an API on aluminium in the range 4–21 μg cm^{-2}.[33] Since the commencement of my research project, there have been publications and research utilising fibre optic grazing-angle FTIR systems where APIs have been detected down to loadings of less than 0.1 μg cm^{-2} on stainless steel and down to 0.33 μg cm^{-2} on aluminium.[34, 35]

1.8 Where to now? – The future of cleaning validation

Cleaning validation in the pharmaceutical industry remains a vital component in quality control systems to prevent drug contamination. The future is likely to bring changes and improvements in many areas, with an overall goal of lowering the cost of cleaning and quality control. There are likely to be improvements in locating the sources of contamination and steps taken to either eliminate, or at least to minimise and control, the contamination from these sources; this could be through better personnel training and procedures, or better equipment. There is likely to be a trend towards automatic cleaning procedures to minimise human error, although there will always be the need for human checks. The design of manufacturing equipment is likely to take cleaning procedures into greater account, with reduction or elimination of hard-to-clean places and implementation of more CIP or closed systems where contamination from external sources is less likely. Cleaning agents are likely to become more environmentally friendly and less toxic, and there may also be improvements to make them more easily detected. Cleaning efficiently should also improve with one product being used throughout an entire plant rather than different agents depending on the nature of the last product produced or surface to be cleaned.[3, 18]

The FDA is unlikely to change its views on sampling and detection methods. Therefore, there should be improvements in the speed of direct sampling methods which analyse the surface directly. Rinse sampling is likely to endure as CIP and closed systems become more common, although such processes may need to be validated with measurements

taken directly from the cleaned surface. The development in analytical methods is likely to be towards direct measurements of the cleaned surface. Analytical methods are likely to become more specific and more sensitive, with results being expected over shorter and shorter time scales. One goal is an inline "cleanliness sensor", which can monitor the cleanliness of a system and give feedback on cleaning performance. Changes to the regulatory bodies may mean that a consensus could be achieved on how to calculate RLs, which could allow for the harmonisation of regulatory requirements and the globalisation of cleaning validation protocols.[3]

In the meantime, the development and adoption of simple analytical techniques that allow the quantification of contamination at the acceptable RLs, with the adoption of less complicated methods which reduce the potential for problems should help to minimise the cost of cleaning.[24]

1.9 References

1. APIC, Cleaning validation in the active pharmaceutical ingredient manufacturing plants, European Chemical Industry Council CEFIC, 1999.
2. Chawla, M. K. In *Surface Contamination and Cleaning*; Mittal, K. L.; VSP: Ridderkerk, Netherlands, 2003; Vol. 1, 23-42.
3. Bismuth, G. and Neumann, S. *Cleaning Validation: A Practical Approach*; Interpharm/CRC Press LLC: Boca Raton, Florida, 2000.
4. Amer, G. and Deshmane, P. Ensuring successful validation: The logical steps to efficient cleaning procedures, *BioPharm.*, **2001**, *14*, 26-32.
5. Food and Drug Administration (FDA), Guide to inspections validation of cleaning processes, Office of Regional Operations, 1993.
6. Dourson, M. L. and Stara, J. F. Regulatory history and experimental support of uncertainty (safety) factors, *Regulatory Toxicology and Pharmacology*, **1983**, *3*, 224-238.

7. Layton, D. W., Mallon, B. J., Rosenblatt, D. H. and Small, M. J. Deriving allowable daily intakes for systemic toxicants lacking chronic toxicity data, *Regulatory Toxicology and Pharmacology*, **1987**, *7*, 96-112.
8. Fourman, G. and Mullen, M. Determining cleaning validation acceptance limits for pharmaceutical manufacturing operations, *Pharmaceutical Technology*, **1993**, *17*, 54-60.
9. The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH), Q3B(R): Impurities in New Drug Products (Revised Guideline), FDA, 2003.
10. The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH), Q2A: Text on Validation of Analytical Procedures, FDA, 1994.
11. The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH), Q2B: Validation of Analytical Procedures: Methodology, FDA, 1996.
12. The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH), Q3C: Impurities: Guideline for Residual Solvents, FDA, 1997.
13. The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH), Q6A: Specifications: Test Procedures and Acceptance Criteria for New Drug Substances and New Drug Products: Chemical Substances, FDA, 1999.
14. The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH), Q7A: Good Manufacturing Practice Guide for Active Pharmaceutical Ingredients, FDA, 2000.
15. The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH), Q3A(R): Impurities in New Drug Substances (Revised Guideline), FDA, 2002.
16. Food Standards Australia New Zealand, Australia New Zealand Food Standards Code, FSANZ, 2002.

17. Alconox, Pharmaceutical Cleaning Validation Method References for Alconox, Inc. Detergents, Alconox, Inc., 2004.
18. Jenkins, K. M. and Vanderwielen, A. J. Cleaning validation: an overall perspective, *Pharmaceutical Technology*, **1994**, *18*, 60-73.
19. LeBlanc, D. A. "Visually clean" as a sole acceptance criterion for cleaning validation protocols, *PDA Journal of Pharmaceutical Science and Technology*, **2002**, *56*, 31-36.
20. Forsyth, R. J., Van Nostrand, V. and Martin, G. P. Visible residue limit for cleaning validation and its potential application in a pharmaceutical research facility, *Pharmaceutical Technology*, **2004**, *28*, 58-71.
21. Forsyth, R. J. and Van Nostrand, V. Application of visible-residue limit for cleaning validation in a pharmaceutical manufacturing facility, *Pharmaceutical Technology*, **2005**, *29*, 152-161.
22. Smith, J. M. Selecting analytical methods to detect residue from cleaning compounds in validated process systems, *Pharmaceutical Technology*, **1993**, *17*, 88-98.
23. LeBlanc, D. A. Sampling, analyzing, and removing surface residues found in pharmaceutical manufacturing equipment, *Microcontamination*, **1993**, 37-40.
24. Kaiser, H. J. In *Surface Contamination and Cleaning*; Mittal, K. L.; VSP: Ridderkerk, Netherlands, 2003; Vol. 1, 75-84.
25. Kaiser, H. J. and Minowitz, M. Analyzing cleaning validation samples: what method?, *Journal of Validation Technology*, **2001**, *7*, 226-236.
26. Forsyth, R. J. and Haynes, D. V. Cleaning validation in a pharmaceutical research facility, *Pharmaceutical Technology*, **1998**, *22*, 104-112.
27. Westman, L. and Karlsson, G. Methods for detecting residues of cleaning agents during cleaning validation, *PDA Journal of Pharmaceutical Science and Technology*, **2000**, *54*, 365-372.
28. Kaiser, H. J., Tirey, J. F. and LeBlanc, D. A. Measurement of organic and inorganic residues recovered from surfaces, *Journal of Validation Technology*, **1999**, *6*, 424-436.

29. Biwald, C. E. and Gavlick, W. K. Use of total organic carbon analysis and Fourier-transform infrared spectroscopy to determine residues of cleaning agents on surfaces, *Journal of AOAC International*, **1997**, *80*, 1078-1083.

30. Griffiths, P. R. and De Haseth, J. A. *Fourier Transform Infrared Spectroscopy*; John Wiley & Sons: New York, 1986.

31. Gavlick, W. K., Ohlemeier, L. A. and Kaiser, H. J. Analytical strategies for cleaning agent residue determination, *Pharmaceutical Technology*, **1995**, *19*, 136-144.

32. Druy, M. A. and Bolduc, R. A. Fiber optic non-contact reflectance probe for detection of contamination in pharmaceutical mixing vessels, *Proceedings of SPIE-The International Society for Optical Engineering*, **1999**, *3538*, 167-171.

33. Mehta, N. K., Goenaga-Polo, J., Hernandez-Rivera, S. P., Hernandez, D., Thomson, M. A. and Melling, P. J. Development of an in situ spectroscopic method for cleaning validation using mid-IR fiber-optics, *BioPharm*, **2002**, *15*, 36-42, 71.

34. Teelucksingh, N. and Reddy, K. B. Quantification of active pharmaceutical ingredients on metal surfaces using a mid-IR grazing-angle fiber optics probe - An in-situ cleaning verification process, *Spectroscopy*, **2005**, *20*, 16-22.

35. Druy, M. A. Applications for mid-IR spectroscopy in the pharmaceutical process environment - Mid-IR continues to prove its usefulness as a process analytical technology, *Spectroscopy*, **2004**, *19*, 60-63.

Chapter 2

Infrared Spectroscopy

2.1 Infrared spectroscopy

The infrared (IR) region of the electromagnetic spectrum covers the range 12500–10 cm^{-1}. The mid-infrared region extents from 4000–400 cm^{-1} and is flanked by the far-infrared region at lower wavenumbers (to the red) and the near-infrared region at higher wavenumbers (to the blue). IR spectroscopic techniques allow spectra to be collected from gases, liquids and solids. They can be easy to use and are rapid, sensitive and non-destructive; therefore it is not surprising that IR spectroscopy is amongst the most commonly used analytical spectroscopic methods.[1]

IR spectroscopy is a well established technique for investigating molecular vibrations. For a molecule to have an IR vibrational spectrum, one or more of the molecular vibrations need to be associated with a change in the electric dipole moment. Therefore homonuclear diatomic molecules, such as H_2 and O_2, do not exhibit IR vibrational spectra. Non-linear polyatomic molecules have IR vibrational spectra, but not all vibrational modes are necessarily active.

The IR spectra of polyatomic molecules can be complex; they contain information about the different types of bonds within a molecule and can be used for structural analysis. The vibrational modes are generally categorised as arising from either stretching or deformation vibrations. Stretching modes tend to display fundamental bands between 4000–1500 cm^{-1} whereas deformation modes tend to be in the range 1500–500 cm^{-1}.[2] The deforming vibrations can be further separated into bending, twisting, wagging and rocking modes. When there are two or more local vibrational coordinates that are equivalent (or near equivalent) by symmetry, their combinations are labelled as symmetric or asymmetric, depending on their phase relationships. The region 1400–900 cm^{-1} of an IR spectrum is sometimes known as the fingerprint

region. It relates mostly to deformation bands, which represent large fragments of the molecule and are characteristic of particular groups of molecules.

In general, there are four ways that incident light can interact with a sample (Figure 2.1). It can be reflected from the sample surface (with intensity I_R), scattered by the surface or bulk of the sample (I_S), absorbed (I_A) or transmitted though the sample (I_T). Conservation of energy requires that the intensities associated with these must sum to the intensity (I_0) incident on the sample (equation (2.1)).

$$I_0 = I_R + I_S + I_A + I_T \tag{2.1}$$

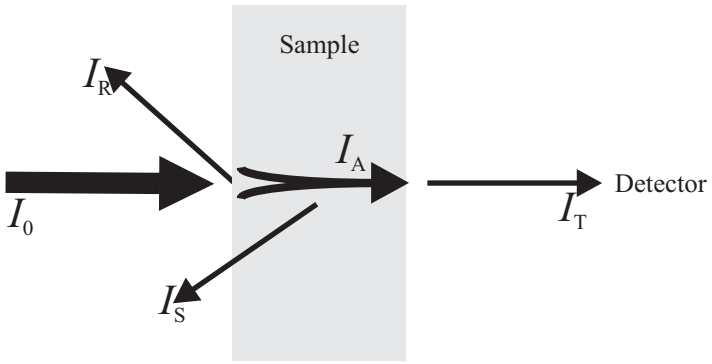


Figure 2.1: Radiation with intensity I_0 incident on a sample can suffer four fates: reflection (with intensity I_R), scattering (with intensity I_S), absorption (with intensity I_A) and transmission (with intensity I_T).

Each of the intensities on the right of equation (2.1) contains information about the sample. The most commonly used is the absorbed intensity, but I_A can not normally be measured directly. Instead, it is inferred by measuring other intensities. In a typical absorption spectrum, only I_0 and I_T are measured and the sum of I_S and I_R is assumed to be negligible. If I_R is measured instead of I_T then a reflection spectrum results.

2.2 Fourier transform infrared spectrometers (FTIR)

In general, there are two types of mid-IR spectrometers; dispersive and Fourier transform (FT). The majority of modern IR spectrometers are FT instruments, which have the advantage that data are collected at many wavelengths simultaneously (rather than consecutively) resulting in faster and more sensitive measurements.

In an FT instrument, spectral elements are recorded simultaneously in the time domain. Since IR detectors are not fast enough to measure the radiation frequencies (10–100 GHz for IR radiation) directly, FTIR spectrometers use a Michelson interferometer to modulate the IR radiation (Figure 2.2). The source radiation is directed through a beam splitter, which produces two beams of equal intensity. These beams are then reflected off mirrors, one fixed and the other movable. The movable mirror allows the adjustment of the path length of one of the beams. When the beams are recombined at the beam splitter, any given wavelength can be in or out of phase (depending on the difference between the two path lengths) giving rise to constructive or destructive interference, respectively. The intensity of each wavelength fluctuates in a predictable manner and the total intensity can be graphed *versus* the change in path length (retardation) to give an interferogram (Figure 2.3). The interferogram is composed of a strong centre burst (corresponding to zero retardation, when all wavelengths are in phase) and decaying wings whose intensities decrease as the distance from the centre burst increases.

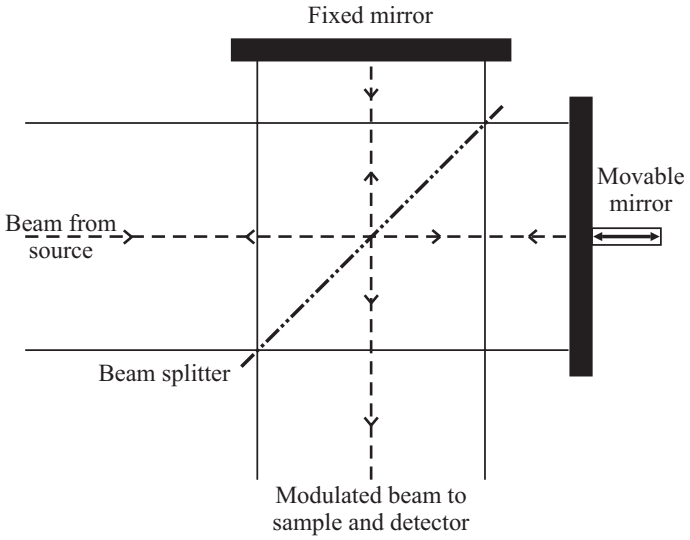


Figure 2.2: Schematic of a Michelson interferometer.

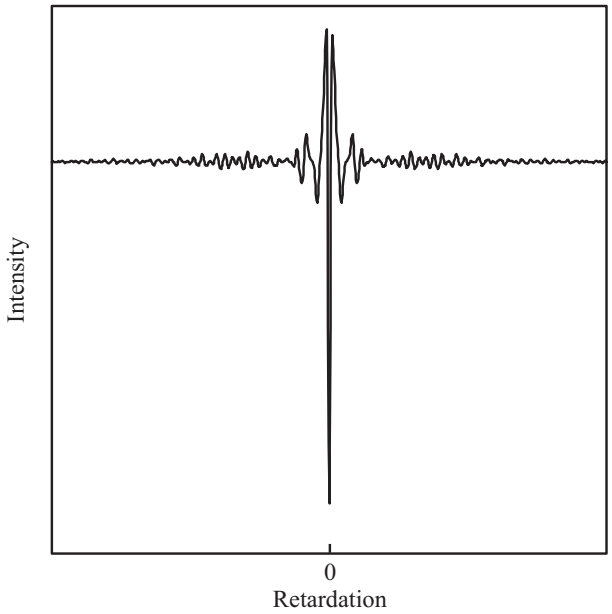


Figure 2.3: Interferogram collected at the detector of an FTIR spectrometer. The decay wings have been truncated so the centre burst structure can be seen.

2.2.1 The interferogram

As will be shown below, an ideal interferogram is symmetrical about the centre burst, so only one half of it is needed to calculate the spectrum (thereby saving time and computing power). An interferogram can become unsymmetrical (or "chirped"*) under a number of circumstances, the most common being misalignment of either the mirrors or laser, or a divergent radiation source.

The effect of a divergent beam is illustrated in Figure 2.4. The outer parts of the beam travel further, causing the radiation to shift out of phase with respect to radiation at the centre of the beam. This affects the shape of the interferogram and hence the quality of the spectra. Generally, the problem is greater at higher wavenumbers since they correspond to shorter wavelengths and therefore suffer greater phase shifts for a given retardation difference.[3] Another effect of divergence is that the resulting beam is wider at the sample. Apertures placed before and/or after the interferometer are commonly used to reduce these problems.

* Chirping refers to the situation when an interferogram does not contain a point of stationary phase; in general all rapid scanning interferograms are slightly unsymmetric and hence chirped. 3. Griffiths, P. R. and de Haseth, J. A. *Fourier Transform Infrared Spectroscopy*; John Wiley & Sons: New York, 1986.

Misaligned mirrors cause a range of pathlengths resulting in phase differences between different parts of the beam. This effect cannot be compensated through the use of apertures, hence a very important part of FTIR spectrometer maintenance is ensuring that the mirrors move smoothly and are not significantly misaligned.

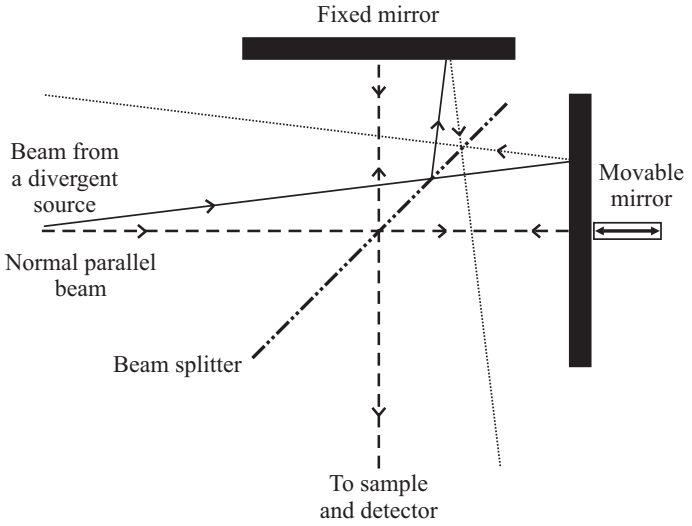


Figure 2.4: Schematic of divergent source radiation passing through a Michelson interferometer showing the change in pathlength that results in a phasing error and loss of resolution in the resulting spectra.

If the reference He-Ne laser of an FTIR instrument is misaligned, a linear wavenumber shift is introduced into the calculated IR spectrum. This is easily corrected by changing the reference wavenumber used in the Fourier transform; but any time the laser is altered the spectrometer should be recalibrated. A similar effect can occur if the aperture at the focus is altered, since this changes the average angle of the beam passing through the spectrometer.

When the recombined radiation is passed though a sample, it is modified by absorption of some wavelengths. The absorption features are embedded in the modified interferogram, but cannot be readily appreciated until the data are transformed to the frequency domain, Figure 2.5. Generally, two “single-channel” (or “single-beam”) interferograms are collected; a “sample” channel with the sample in place and a “background” or “reference channel”, without the sample in place. The final FTIR spectrum is calculated from the two channels as described below.

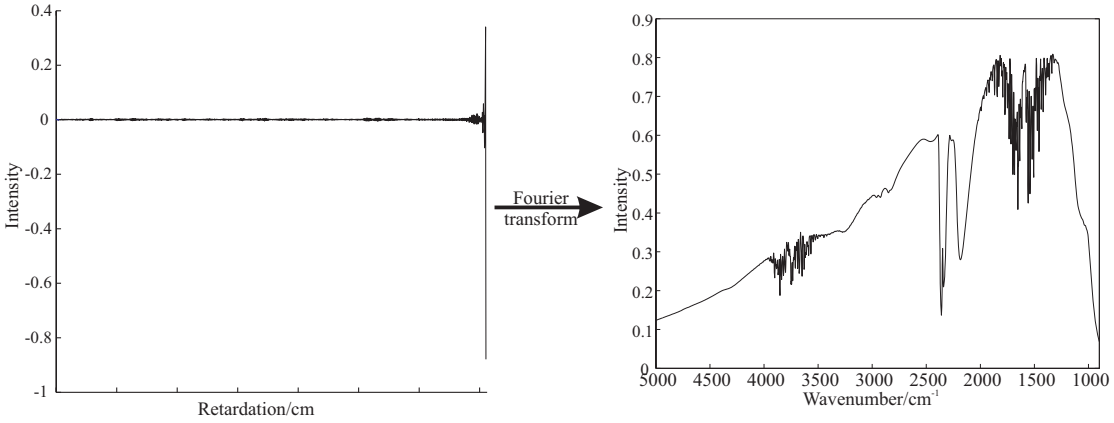


Figure 2.5: The plot on the left shows half of the interferogram, the IR spectrum in the time domain. The interferogram is converted by Fourier transform into the frequency domain, as displayed on the right. The bands shown in the single-channel spectrum are from water vapour (the two branched bands with fine structure), carbon dioxide at ~2400 cm^{-1} and H-Se band (due to absorption by the IR fibre optic) at ~2200 cm^{-1}. Detector non-linearities cause the apparent variation of the baseline absorbance with wavenumber.

The following mathematical description of the FT processes is based on Griffiths and de Haseth's treatment in *Fourier Transform Infrared Spectroscopy*.[3] It assumes a perfectly collimated beam and, initially, the beam splitter is taken to be ideal in that it is non-absorbing and has reflectance and transmittance of exactly 50% each.

The optical path difference between the fixed and movable mirrors is referred to as retardation and has the symbol δ. The intensity of the source radiation at wavenumber $\bar{\nu}$ is $I_0(\bar{\nu})$. The retardation-dependent intensity of the radiation measured by the detector at the same wavenumber is $I'(\delta, \bar{\nu})$, which is given by equation (2.2).

$$I'(\delta, \bar{\nu}) = 0.5\, I_0(\bar{\nu})\left(1+ \cos 2\pi\bar{\nu}\delta\right)$$

$$= 0.5\, I_0(\bar{\nu}) + 0.5\, I_0(\bar{\nu}) \cos 2\pi\bar{\nu}\delta \qquad (2.2)$$

The second, modulated component of equation (2.2) contains the spectroscopic information and contributes to the interferogram. It can be corrected for a non-ideal beam splitter, as well as non-linear detector and amplifier responses, by the inclusion of a wavenumber-dependent function, $H(\bar{\nu})$, to give:

$$I(\delta, \bar{\nu}) = 0.5\, I_0(\bar{\nu})\, H(\bar{\nu}) \cos 2\pi\bar{\nu}\delta \qquad (2.3)$$

The factors outside the cosine of equation (2.3) can be combined into a single intensity factor, $B(\bar{\nu})$, to give equation (2.4).

$$I(\delta, \bar{\nu}) = B(\bar{\nu}) \cos 2\pi\bar{\nu}\delta \qquad (2.4)$$

Since most Michelson interferometers scan the movable mirror at a constant velocity, υ, the signal collected at the detector can be expressed as a function of time rather than retardation. The relationship between time and retardation is shown in equation (2.5).

$$\delta = 2\upsilon t \tag{2.5}$$

This can be substituted back into equation (2.4) to give an expression for the interferogram in terms of time rather than retardation.

$$I(t, \bar{\nu}) = B(\bar{\nu})\cos 4\pi\bar{\nu}\upsilon t \tag{2.6}$$

The abscissa units used in the interferogram (equation (2.4) or (2.6)) determine those for the resulting spectrum. If retardation (in cm) is used, the spectrum will be expressed in cm^{-1} whereas if the interferogram is expressed in terms of time (in seconds) then the spectral units will be Hertz.

Equation (2.7) shows that, assuming the ideal situation outlined earlier, the interferogram collected from a monochromatic radiation source would be a cosine function. The general form for any time-dependent cosine wave of frequency f and maximum amplitude A_0 after time t is equal to $A(t)$.

$$A(t) = A_0 \cos 2\pi f t \tag{2.7}$$

Consequently, the interferogram frequency corresponding to radiation of wavenumber $\bar{\nu}$ is proportional to scanning velocity and given by;

$$f_{\bar{\nu}} = 2\upsilon\bar{\nu} \tag{2.8}$$

The expressions for interferograms can be extended to polychromatic radiation sources by integrating over all wavenumbers, as shown in equation (2.9).

$$I(\delta) = \int_{-\infty}^{+\infty} B(\bar{\nu})\cos 2\pi\bar{\nu}\delta \, d\bar{\nu} \tag{2.9}$$

$I(\delta)$ can be described mathematically as being the cosine Fourier transform of $B(\bar{\nu})$. The complementary transform function is shown in equation (2.10).

$$B(\bar{\nu}) = \int_{-\infty}^{+\infty} I(\delta)\cos 2\pi\bar{\nu}\delta \, d\delta \tag{2.10}$$

Since cosines are even functions about zero, the last two integrals can be rewritten as;

$$I(\delta) = 2\int_0^{+\infty} B(\bar{\nu})\cos 2\pi\bar{\nu}\delta \, d\bar{\nu} \tag{2.11}$$

$$B(\bar{\nu}) = 2\int_0^{+\infty} I(\delta)\cos 2\pi\bar{\nu}\delta \, d\delta \tag{2.12}$$

2.2.2 Spectral resolution

The spectral resolution of an FTIR instrument is mainly determined by the maximum retardation;[1, 4] the bigger the range of pathlength difference the better the resolution. If it were possible to carry the integrals in equations (2.11) and (2.12) over the full range, from 0 to ∞, from an interferogram digitised at infinitesimally small intervals, a spectrum of infinitely high resolution could be obtained, in principle. These requirements are obviously not possible, and the resolution of a spectrum is limited by constraints on the spectrometer and computing power.[3]

In order to understand the practical requirements of instrumental resolution, consider two closely spaced signals separated by $\Delta\bar{\nu}$:

$$\Delta\bar{\nu} = \bar{\nu}_1 - \bar{\nu}_2 \tag{2.13}$$

The interferogram for the two signals will consist of two cosine waves with slightly different frequencies. These waves will be in phase at zero retardation, completely out of phase at $\delta = 0.5/\Delta\bar{\nu}$ and then back in phase again at $\delta = 1/\Delta\bar{\nu}$. To distinguish between the two signals, it would seem probable that the interferogram should be collected over at least one full phase cycle, which corresponds to a minimum retardation $1/\Delta\bar{\nu}$. This would suggest the best resolution that can be achieved by an interferometer with maximum retardation δ_{max} is:

$$\Delta\bar{\nu} = \frac{1}{\delta_{max}} \tag{2.14}$$

In fact, equation (2.14) is slightly pessimistic. If one accepts that two signals are sufficiently well resolved when there is ~20 % dip between them (the Rayleigh criterion),[3] a more realistic estimation is $\Delta\bar{\nu} = 0.75/\delta_{max}$.

2.2.3 Instrument line shape and apodisation

Many aspects of the spectrometer, including its components (mirrors, beam splitter, detector, system electronics, etc.) and sampling methods, have a bearing on the interferogram and hence on the spectrum and its resolution. These contributions collectively determine the "instrument line shape" or "instrument function".[3] One of the factors that can influence the instrument line shape function is the way in which the interferogram is truncated to accommodate the finite maximum retardation of the spectrometer.

The function that is used to truncate the interferogram is called the apodisation function. If the apodisation function is a "box car" (corresponding to rectangular truncation – see Figure 2.6), the instrument line shape function has undulating side lobes. Other apodisation functions, such as trapezoidal and triangular functions, do a better job of suppressing the side lobes, and a Gaussian apodisation (shown in the lower portion of Figure 2.6) does an excellent job.

There is a cost to side-lobe suppression in that the spectral resolution is decreased. This effect can be seen in Figure 2.6, where the central band for the box-car function is narrower than that for the Gaussian function. In general functions that do the worst job of suppressing the side lobes produce spectra that have the highest resolution.

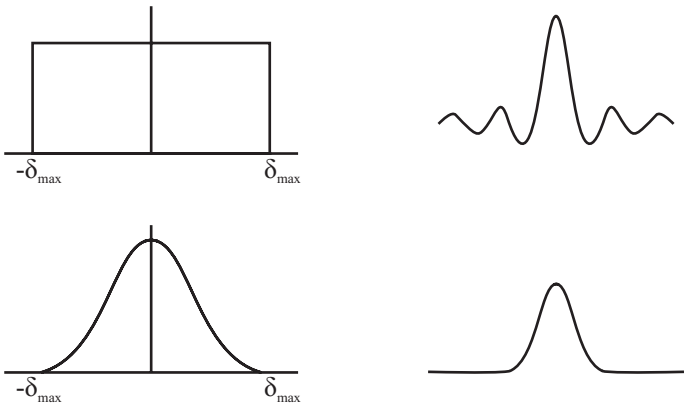


Figure 2.6: Left hand column shows the shape of the apodisation function which is used to truncate the interferogram; top is a box-car function and bottom is a Gaussian function. The right hand column shows the resulting single-band spectra when the Fourier transform is performed on the apodised interferogram.

2.3 Detectors

The most common detector in FTIR instruments is a deuterated triglycine sulfate (DTGS) detector. In situations where DTGS detectors are inappropriate, they are usually replaced by mercury cadmium telluride (MCT) detectors.

DTGS detectors are widely used since they have a broad working range (400–4000 cm^{-1}) with a good linear response. They are simple to use, robust, inexpensive and can be operated at room temperature. The principle behind their operation is a change in the dielectric constant, and therefore the capacitance, of the DTGS material induced by changes in the temperature on absorption of IR radiation. The signal is measured as a voltage across the detector. The major disadvantage of the DTGS detector is a long response time; high spectral resolution requires a slow mirror velocity and hence fewer interferograms can be collected in a set time period than for quicker detectors. The interferometer scan time required to achieve a spectrum with a resolution of 4 cm^{-1} when using a DTGS detector is around 1 second.[1]

An MCT detector works through an internal photo-effect. The mercury, cadmium and telluride alloy is a semiconductor with a bandgap that corresponds to the energy of IR photons. When IR radiation falls on the detector, the electrons in the valence band are excited into the conduction band and can therefore respond to an applied voltage to give an electrical current. The high wavenumber end of their spectral range is ~4000 cm^{-1}. The low wavenumber limit is determined, according to the bandgap, to be between ~400 and 900 cm^{-1} and the "cut-off" is sharper than at the high wavenumber limit. Detectors with a lower bandgap (and therefore able to detect lower energy radiation) are more susceptible to thermal excitation of electrons into the conduction band and are therefore noisier at a given temperature. Consequently, they require cooling, usually achieved by using liquid nitrogen, which requires mounting in a vacuum Dewar and makes them generally more expensive than DTGS detectors.[4]

The advantages of an MCT detector relate to response time and sensitivity. Response is about 100 times faster than for a DTGS detector,[1] meaning more scans can be collected and averaged in a given period. This has advantages in the signal-to-noise

ratio and also means that rapidly changing systems can be studied over a shorter time scale. MCT detectors are also very sensitive, although this does mean that they can be easily saturated. Another limitation is that, in many cases, response is not linear over the entire spectral range, although this can be compensated through electronics and software.[1]

2.4 Surface analysis using infrared spectroscopy

The properties of a thin film deposited on a metal or dielectric surface can be very different from those of the bulk phase. These differences can have important consequences in fields such as catalysis, where surface-adsorbed species can greatly increase the speed and selectivity of a particular reaction. Of more relevance to this study, a thin film can have significantly modified optical properties.

As mentioned in Section 2.1 , there are a number of ways that an infrared spectrum can be collected from a sample, depending on the sample's properties and the information required. To analyse materials that have been deposited on a surface, one can use absorption, diffuse reflectance, photo-acoustic, attenuated total reflectance, emission or reflection absorption spectroscopy. IR reflection absorption spectroscopy (IRRAS; known less commonly as reflection absorption infrared spectroscopy (RAIRS) or infrared external reflection spectroscopy (IR-ERS)) is the most suitable for the systems of interest in this work, where chemical residues, such as pharmaceuticals or surfactants, have been deposited on relatively clean metallic or dielectric surfaces. Grazing-angle FT-IRRAS allows an analyte spectrum to be measured directly from a sample surface. It is also capable of sampling relatively large areas, which provides the experimentalist with a good idea of the average surface loading.

Although the spectroscopy of surface adsorbed species has been of interest for a long time, prior to the 1960s IR studies of molecules adsorbed on non-transparent substrates were difficult.[5] The problem was principally one of achieving a long enough path length through the film to obtain sufficient absorption. In one attempt to get around this, the molecules of interest were adsorbed onto small metal particles,

which were supported on a metal oxide substrate. Unfortunately, it was discovered that interactions between the oxide and the metal meant that the spectrum was not the same as when a sample is adsorbed onto a bulk metal substrate.[6] Another technique used IR radiation at an incidence angle of 0° (normal to the surface) where the radiation was reflected back to the detector.[6] For strong reflection, the support had to be a metallic mirror, but the spectra were poor since the radiation produces a standing wave with a node at the surface (see below). The consequence of this is that the electric field vector at the substrate surface is approximately zero and there is little absorption. Attempts to use a multi-pass cell arrangement utilising two mirrors did not improve the signal enough to provide an adequate spectrum.[6, 7]

IRRAS was pioneered by researchers such as Greenler[6] and Francis and Ellison,[8] who found that it could be used to investigate thin films and even monolayers adsorbed on metal substrates.[9] Figure 2.7 shows a schematic of a thin film supported on a quasi-infinite support substrate, defining some of the parameters important to IRRAS. The film thickness is l and the real refractive indices for the media are denoted n_1, n_2 and n_3 for air, the film and the substrate, respectively, (it should be noted that the refractive indices are dependent on the radiation wavelengths). When the sample and substrate substances are absorbing (rather than transparent) the refractive indices are complex, $\hat{n}_j = n_j + ik$ where k is the absorption index,[†] and θ_1, θ_2 and θ_3 are the angles, relative to the surface normal, at which the radiation passes through air, the film and the substrate, respectively. θ_1 is usually selected by the experimentalist and is known as the incidence angle.

Many of the initial studies using the IRRAS techniques focused on multiple reflections at large incidence angles,[11, 12] with the view that this would permit a reasonable signal to be collected. Since peak height was not proportional to number of reflections, much work was performed on determining the ideal number, which was found to be dependent on the substrate, film and film thickness.[11] With the advances of technology offered by FTIR spectrometers, stronger IR sources and more sensitive

[†] Some other texts use the convention $\hat{n} = n - ik$; k describes the amount of absorption of radiation at particular wavelength as it passes through the relevant medium. It is related to the linear absorption coefficient α *via* the expression $\alpha = 4\pi\tilde{\upsilon}k$ where $\tilde{\upsilon}$ is the wavenumber and k is the absorption index.10. Bertie, J. E. In *Handbook of Vibrational Spectroscopy: Theory and Instrumentation.*; Chalmers, J. M. and Griffith, P. R.; John Wiley and Sons, LTD: Chichester, UK, 2002; Vol. 1, 88-100.

detectors, multiple reflection techniques are no longer required and most IRRAS spectra are now collected using only a single reflection.

Of the radiation incident on the interface between two media with different refractive indices, a part undergoes refraction, while some is reflected. In a system with two (or more) interfaces, the light detected after leaving the primary surface (the upper surface of Figure 2.7) will be a result of various orders of reflection and refraction from all interfaces. If the media through which the radiation passes also absorb radiation, this fact will further modify the detected signal and the resulting spectrum is then known as reflectance-absorbance since it is a combination of reflected and refracted radiation.

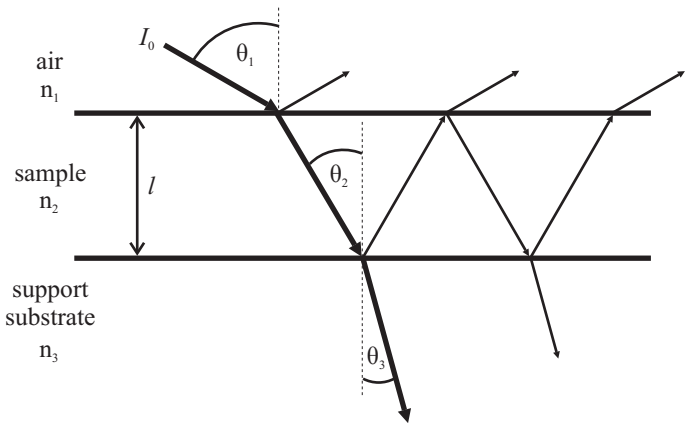


Figure 2.7: Schematic of a light beam passing through a sample adsorbed on a support substrate. See the text for a description of the parameters.

2.4.1 Polarisation of light

The interaction between light and the molecules present in a medium depends, in part, on orientation of the electric dipole transition moments of the molecule with respect to the electric field vector of the radiation.[‡] In the wave model of electromagnetic radiation, unpolarised light is associated with electric-field waves whose electric vectors are oriented in random directions perpendicular to the direction of propagation. Plane polarised light describes the situation where the electric vectors of all component waves are parallel to each other. In surface spectroscopy, there are two special types of plane polarisation; p-polarised light has its electric vector in the plane

[‡] It is assumed here that only electric-dipole transitions are important.

defined by the incident and reflected rays (and therefore has a component perpendicular to the surface) whereas s-polarised light has its electric vector parallel to the surface, Figure 2.8.

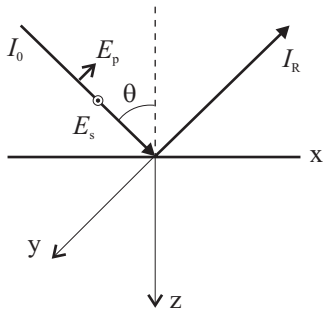


Figure 2.8: An incident light beam of intensity I_0, propagating in the plane of the page (x-z plane), is incident, at an angle θ on a surface that is perpendicular to the page (x-y plane), and reflected with intensity I_R. The electric field vectors for the p- and s-polarised components of the light are shown as E_p and E_s respectively. The p-polarised electric field is in the plane of the incident and reflected rays whereas the s-polarised electric field, E_s, is parallel to the plane of the surface.

In comparison with a transmission IR spectrum of the bulk analyte, the IRRA spectral intensities, peak positions and shapes can be altered. Some of these changes may be due to modified structure of the analyte, orientational effects or chemical bonding of analyte molecules to the substrate. The absence of vibrational features may indicate that the corresponding functional group is strongly modified, for example by being chemisorbed to the surface.[13] But an alternative possibility is that the compound is oriented on the surface in such a way that the vibration can not be excited by the incident radiation.[14] As discussed above, the electric field vector at the substrate surface has a direction defined by the polarisation of the incident radiation. For molecules on the surface to absorb this radiation, the electric field vector must be correctly orientated with respect to the transition moment.[13, 15] The use of polarised radiation therefore makes it possible to study the orientations of anisotropically adsorbed molecules by comparing IRRA band intensities with those measured in transmission or calculated,[13, 16-18] with the proviso that the modes under investigation are not strongly perturbed by interactions with the surface.[13]

2.4.2 Reflection Coefficients

Even when orientational effects and specific analyte–surface interactions are not important, other optical effects can be responsible for spectral changes. These optical

effects are incorporated in the Fresnel coefficients. Differences between transmission and IRRA spectra can be due to changes in reflectivity brought on by rapid changes in the refractive index around an absorption band.[19] The types of "distortions" that can be induced include peak shifts, band-shape changes and, in some cases, apparent band splittings. These effects are dependent on the film thickness and refractive index, as well as the wavelength, incidence angle and polarisation of the incident radiation.[9, 13, 19] In general, major distortions (such as band splitting) are seen only in the spectra of thick films of analytes with very strong absorption features, for example Cu_2O ($k_{max} \approx 3.4$) on copper.[9, 19] For organic compounds, where $n \approx 1.5$ and $k \approx 0.1$, such distortions are less prominent and are largely limited to peak shifts of several cm^{-1}. The breadth of the bands can also influence the shifts, with narrow bands being shifted less than broad ones. Another effect that can influence the peak shape and position is the change in reflectivity between the system with and without the analyte film.[19]

An incidence beam of any general polarization can be considered as a superposition of p- and s-polarised light. For each of these, the reflectance, R_ρ (ρ = p or s), can be determined as the square of the relevant Fresnel coefficient, r_ρ.[20-22]

$$R_\rho = r_\rho^2 \tag{2.15}$$

Where the Fresnel coefficients can be determined according to:

$$r_p = \frac{(n_2\cos\theta_1 - n_1\cos\theta_2)}{(n_2\cos\theta_1 + n_1\cos\theta_2)} \tag{2.16}$$

$$r_s = \frac{(n_1\cos\theta_1 - n_2\cos\theta_2)}{(n_1\cos\theta_1 + n_2\cos\theta_2)} \tag{2.17}$$

The incidence angle, θ_1, is set in the experiment and the other angles can be determined using Snell's law (equation (2.18)). To calculate θ_2,

$$n_1\sin\theta_1 = n_2\sin\theta_2 \tag{2.18}$$

$$\cos\theta_2 = \sqrt{1 - \left(\frac{n_1}{n_2}\right)^2 (1 - \cos^2\theta_1)} \tag{2.19}$$

Determination of θ_3 proceeds in the same way with appropriate substitutions of parameters, and if any of the media are infrared absorbing, their complex refractive indices should be used. The reflectance-absorbance (RA) spectrum can be determined from the reflectances R and R_0, in the presence and absence of a surface film, respectively.

$$RA = -\log_{10}\left(\frac{R}{R_0}\right) \tag{2.20}$$

2.4.3 Metal surfaces

The IR refractive indices of metals are generally quite similar, being of a comparable magnitude and with simple dispersion, changing monotonically and relatively slowly, over the IR region.[12] The case of aluminium is shown in the left plot of Figure 2.9 since it is typical of most metals and is a surface that is used in work reported in this thesis.

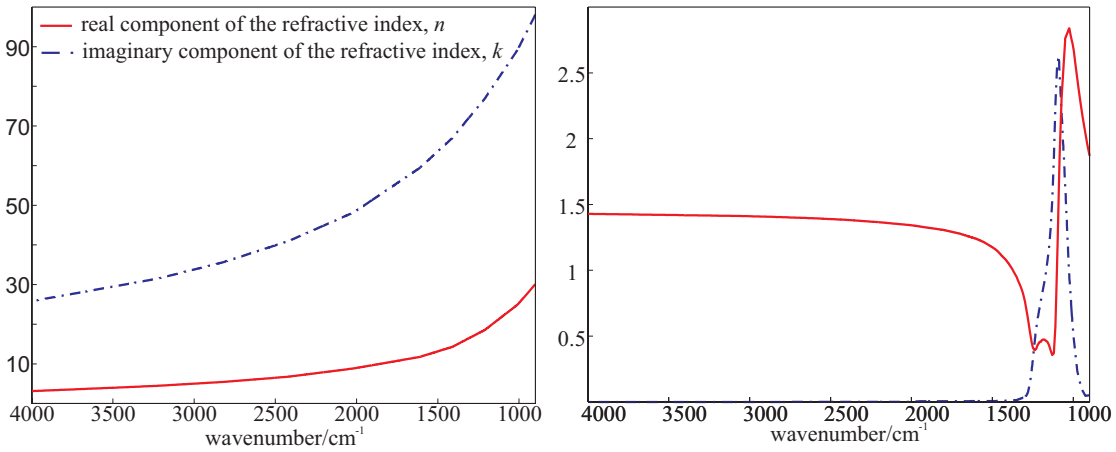


Figure 2.9: Left: real and imaginary components of the refractive index for aluminium over the mid infrared region. Right: real and imaginary components of the refractive index for glass over the mid infrared region.

From the refractive index data, the Frensnel coefficients, r_p and r_s, can been determined and expressed in either Cartesian coordinates (with real and imaginary parts) or polar coordinates (with radial and angular parts). When using polar coordinates, the angle describes a change in the phase of the radiation that occurs on reflection from the corresponding interface. Such changes depend on the polarization of the radiation and the incidence angle as shown in Figure 2.10, where the refractive

indices are n_1 = 1 (for air) and $\hat{n}_3$ = 9 + 49i, which corresponds to aluminium at 2000 cm^{-1} (see left panel of Figure 2.9).

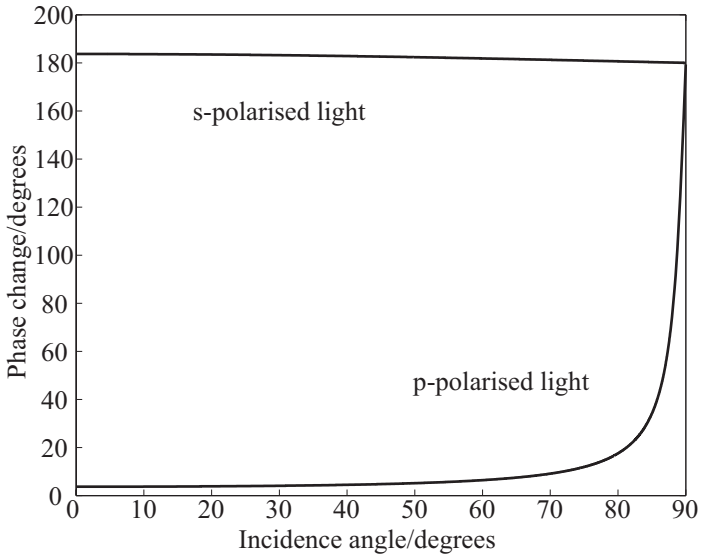


Figure 2.10: The phase change on reflection from a metal mirror substrate is dependent on the angle of the incidence radiation and on the polarisation of the light.[6] The refractive indices are n_1 = 1 (for air) and $\hat{n}_3$ = 9 + 49i, which corresponds to aluminium at 2000 cm^{-1}.

Figure 2.10 shows that the phase change on reflection from a metal surface is approximately 180° for s-polarised light, almost independent of incidence angle. This means that the amplitudes of the incident and reflected s-polarised waves almost cancel at the surface, so the resultant electric field vector is also almost zero. On the other hand, the phase change of p-polarised radiation is strongly dependent on the incidence angle of the incoming radiation. As the incidence angle approaches 90°, the phase change tends to 180° since p-polarized light at θ_1 = 90° is actually s-polarized. But at lower incidence angles the phase change is smaller and the electric field vector at the surface has an appreciable component perpendicular to the surface.[6]

The higher the amplitude of the electric field vector at the surface, the stronger will be its interactions with (appropriately oriented) molecules in the surface film and hence the greater will be the modification of the intensity of "reflected" light. Consequently, the magnitude of the RA for a material deposited on a metal surface will depend on the polarization of the light, the refractive index of the metal and the incident angle. Using equation (2.20), the RA profile over a range of incidence angles can be calculated for each polarisation; Figure 2.11 shows the results for s-, p- and

unpolarised radiation at incidence angles from 0° to 90°. The refractive indices for air and the substrate are the same as for Figure 2.10, while that for the film is $\hat{n}_2 = 1.5 + 0.1i$, which is representative of a moderately strong IR absorption band of an organic molecule. The film thickness is 10 nm and the wavenumber of 2000 cm^{-1} corresponds to the middle of the mid-IR region.

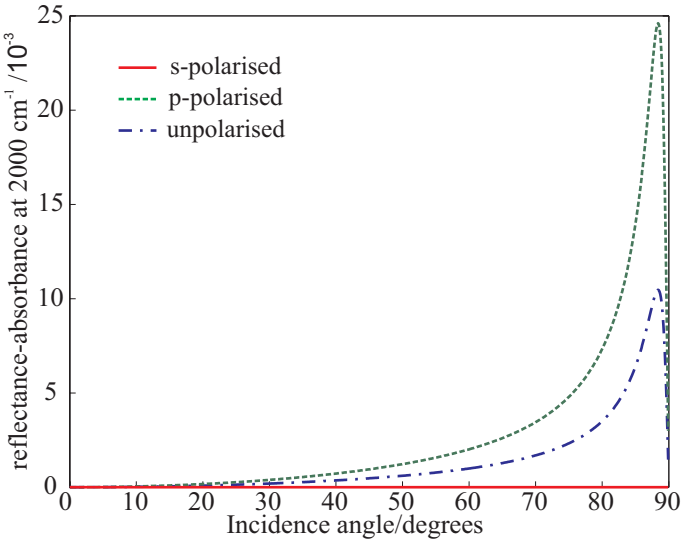


Figure 2.11: Plot of spectral reflection absorption for p-, s- and unpolarised light for a thin (l = 10 nm) film system of the type shown in Figure 2.7. The refractive indices are $n_1 = 1$ (for air), $\hat{n}_2 = 1.5 + 0.1i$ and $\hat{n}_3 = 9 + 49i$.

Figure 2.11 clearly shows that s-polarised light gives almost no RA for all incidence angles, as expected from the electric-field considerations above, (it should also be noted that this is independent of wavelength and film thickness). The RA spectrum for p-polarised radiation is strongly affected by incidence angle, with the strongest signal in Figure 2.11 occurring when the incidence angle is ~88 degrees. Since unpolarised light is a complete mixture of p- and s-polarised components, its RA profile is the average of those for the two components. And since the RA of s-polarised light is essentially zero, the unpolarised RA profile has the same shape as, but only half the magnitude of, that for p-polarisation.

The influence of film thickness on RA is shown in Figure 2.12 for the same system as in Figure 2.11. As expected, the RA increases with thickness. But close inspection of the left plot in Figure 2.12 also shows that the RA maximum moves slightly to lower incidence angles; this becomes more apparent for thicker films, as shown on the right

of Figure 2.12 where the incidence angle for the maximum RA is plotted against film thickness.

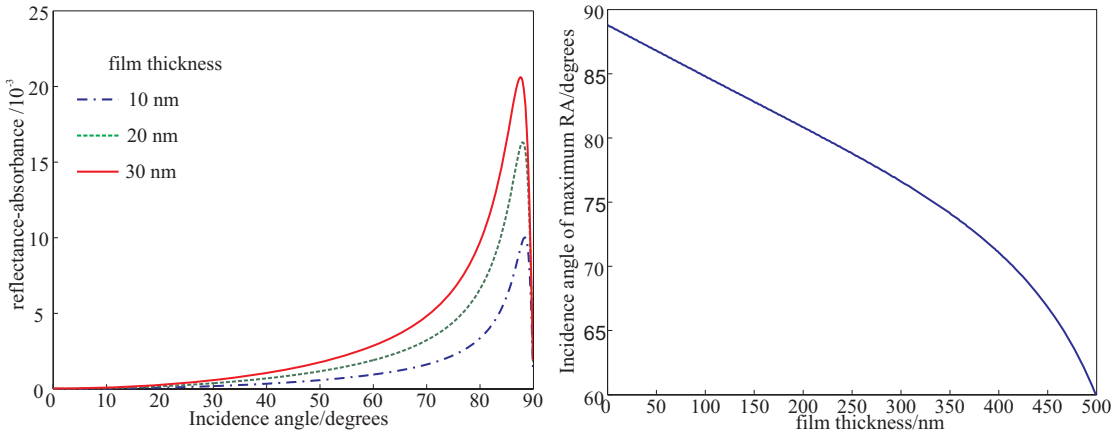


Figure 2.12: Left: the change in reflectance-absorption for unpolarised radiation with incidence angle for systems with increasing film thickness. The calculations were performed at a wavenumber of 2000 cm^{-1} with refractive indices for the film and substrate of $\hat{n}_2$ = 1.5 + 0.1i and $\hat{n}_3$ = 9 + 49i respectively. Right: plot for the same system showing the change of the incidence angle of maximum RA with film thickness.

The RA dependence on wavenumber is shown in the left plot in Figure 2.13. The very simple, near linear, behaviour is a direct result of the simple form of the refractive index shown in Figure 2.9.

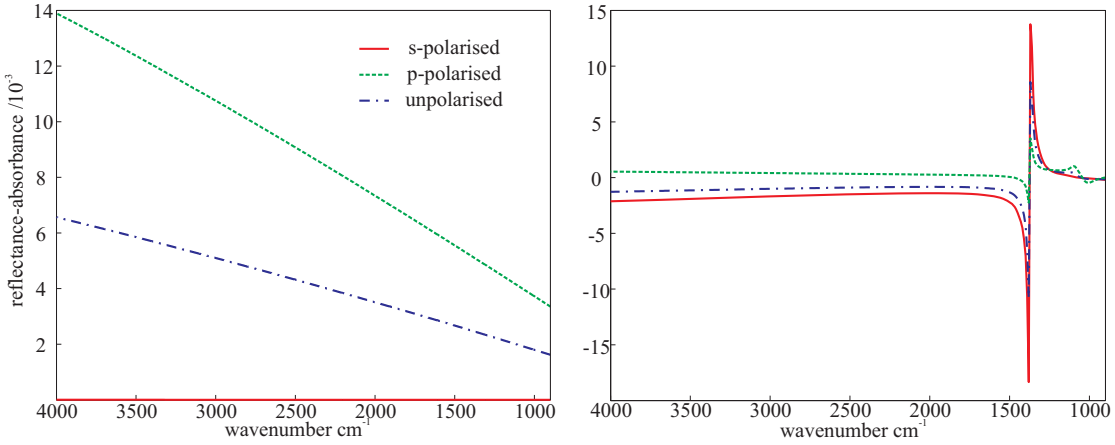


Figure 2.13: Reflection-absorption signal for an organic film ($\hat{n}_2$ = 1.5 + 0.1i), with a thickness 10 nm at incidence angle of 80°. Left: on an aluminium substrate. Right: on glass. The refractive indices used for the two substrates are shown in Figure 2.9.

The conclusions for metallic substrates are that: to a good approximation, only p-polarised radiation needs to be considered in determining optimum conditions; and,

for thin ($\lesssim$ 200 nm) films of organic materials, the optimal incidence angle is ~80–90° (Figure 2.12).

2.4.4 Glass surfaces

The refractive index of silica glass is much more complicated than for metals, as shown in the right-hand panel of Figure 2.9. At wavenumbers greater than ~1500 cm^{-1} glass is transparent to infrared radiation ($k \approx 0$), but at longer wavelengths there are strong absorption bands that are associated with a rapidly fluctuating real part of the refractive index. Since the refractive index changes so drastically over the mid infrared region, the wavelength dependence of the RA is more complicated than for metals (compare the panels of Figure 2.13) and there is no single plot equivalent to Figure 2.11 that is representative of glass over the whole mid infrared region.

Figure 2.13 shows how the RA for glass and aluminium vary with wavenumber at an incidence angle of 80°. Clearly, the behaviour in the case of the metal substrate is much simpler. An important difference relates to the s-polarised spectra. For the metal, there is essentially no s-polarized RA because of destructive interference between the incident and reflected radiation. But for glass, this is not the case since the substrate surface is less reflective and quite transparent in at least some regions of the IR. Another point of difference is the sigmoidal behaviour of the RA at ~1300 cm^{-1} in the glass spectrum, which is due to the rapid changes in refractive index over the same range.

For several reasons, the determination of an optimal incidence angle for a glass substrate is more complicated than for a metal.[14] Not only does the refractive index fluctuate, but also the optical responses to p- and s-polarised radiations can be quite different. For example, at incidence angles below the Brewster angle (the unique incidence angle at which all p-polarised radiation is reflected; ~55° for glass), the signs of the RAs for the two polarisations are the same, but at greater angles they are different.[15]

Blaudez *et al* have investigated a number of parameters that have been suggested as a way to determine the optimal incidence.[15] One method is to maximise the reflectivity difference

$$\Delta R_\rho = R_\rho(l) - R_\rho(0) \tag{2.21}$$

where $R_\rho(l)$ is the RA of ρ-polarised radiation for a substrate coated with a film of thickness l. Other methods include maximizing the normalised reflectivity $\left(\frac{R_\rho(l)}{R_\rho(0)} - 1\right)$ and the signal to noise ratio.[15] From the normalised reflectivity, they suggest that the optimal incidence angle is ~50° for p-polarised radiation and ~20° for s-polarised radiation. For difference reflectivity, the optimal angle is similar for both polarisations; ~80° for p-polarisation and ~75° for s-polarisation. The signal to noise ratio (which has been shown in the referenced literature to be independent of the treatment of the spectral data) gives an optimal incidence angle for s-polarisation of 60°–75°; but the results for p-polarisation are more complicated with different values for positive (~35°) and for negative bands (~75°), although better signal to noise ratios are achieved with the negative bands at the larger angle.[15]

These considerations might seem to indicate that, overall, large incidence angles are better. But Blaudez *et al* considered only polarised radiation. For unpolarised radiation at incidence angles greater than the Brewster angle, the signs of the s- and p-polarised RA spectra are different, and therefore cancel, decreasing the signal to noise ratio. The magnitude of this effect is dependent on the refractive index of the substrate in the region of the analyte bands of interest. That being said, the general conclusion from a more complete study is that large incidence angles are generally to be preferred with unpolarised radiation.[23]

A final consideration when determining the optimal incidence angle is that larger angles give rise to larger IR footprints on the sample surface. This can be a problem for small samples that may not occupy enough of the footprint area to give good spectra. In practical application, this issue has meant that incidence angles larger than ~80° are usually not appropriate. An incidence angle of 80° was employed for the experimental work reported in this thesis.

2.4.5 Infrared reflection-absorption spectra

In the considerations above for metal and glass substrates, it was convenient to assume that $\hat{n}_2$, the complex refractive index of the film, is independent of wavelength. This is, of course, untrue, and it is the dispersion of the refractive index that leads to the IRRA spectrum of the analyte. Generally, there will be wavelength regions where the film is transparent and the real part of the refractive index is fairly constant. There will be other regions that contain absorption bands, which are accompanied by characteristic sigmodial features in the real part of the refractive index. Figure 2.14 shows the real and imaginary part of the refractive index for benzene over the mid IR region as an example.

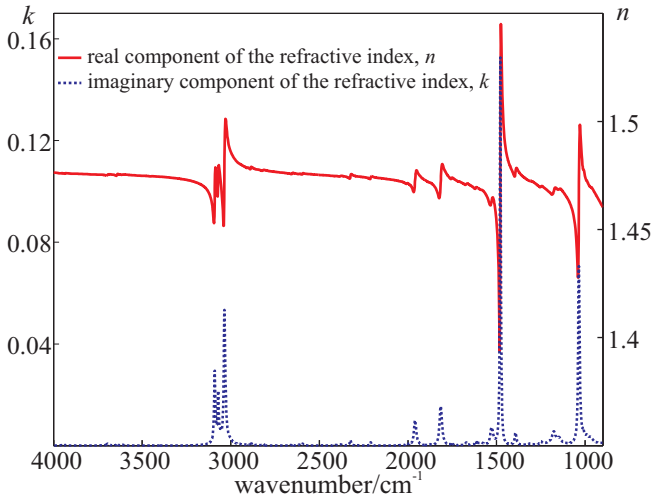


Figure 2.14: Real and imaginary components of the refractive index, for benzene over the mid infrared region.

If the refractive indices of both the film and the substrate are known, then the overall IRRA spectrum can be calculated from them. Figure 2.15 shows the IRRA spectrum for benzene on aluminium determined from the refractive index data in Figures 2.9 and 2.14 at an incidence angle of 80° and a benzene film thickness of 10 nm. It can be seen that the most intense spectrum is obtained with p-polarised radiation. However, an adequate spectrum can still be seen using unpolarised radiation, especially when the inevitable losses incurred by using a polariser are factored in to considerations. The sloping baselines are due to a film-induced phase shift as can be demonstrated by calculating the corrections from the real part of the refractive index, indicated in

Figure 2.15 as dashed lines. In practical applications, the sloping baselines seen here are small in comparison with other baseline effects due to things such as surface roughness and non-parallel surfaces.

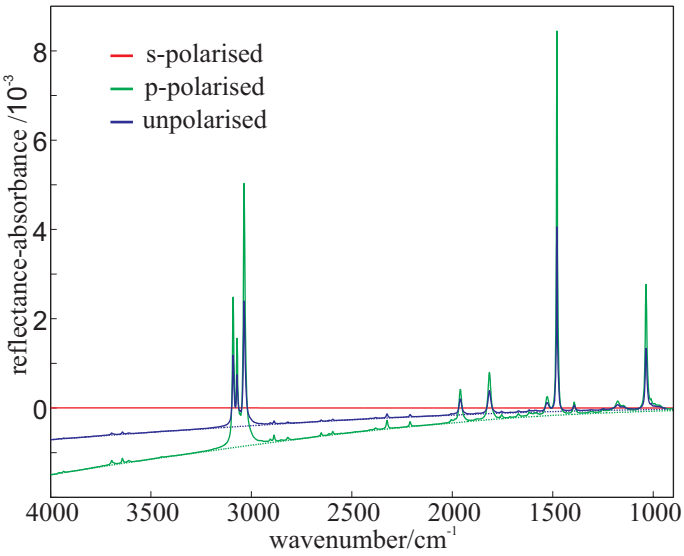


Figure 2.15: IRRA spectra for benzene on aluminium. The s-polarised spectrum has been included to show the zero RA baseline. Shown in dashed lines for p- and unpolarised radiation are the baselines for the theoretical case where the benzene film is non-absorbing.

The equivalent IRRA spectra for benzene on glass (at an incidence angle of 80° and film thickness of 10 nm), calculated using the data in Figures 2.9 and 2.14, is shown in Figure 2.16. The most obvious difference in comparison with the benzene/aluminium system is the presence of very intense (note the difference in ordinate scales between Figures 2.15 and 2.16) glass features at wavenumbers less than 1500 cm^{-1}. These dominate the benzene features and make it difficult to analyse bands that are in close proximity. Fortunately, glass has no significant absorption bands at wavenumbers greater than 1500 cm^{-1}, which means that analyte bands in the region 4000–1500 cm^{-1} can be readily analysed. This is an important factor to be considered when using dielectric substrates. If there are no analyte bands in an absorption window of the substrate, analysis becomes very difficult.

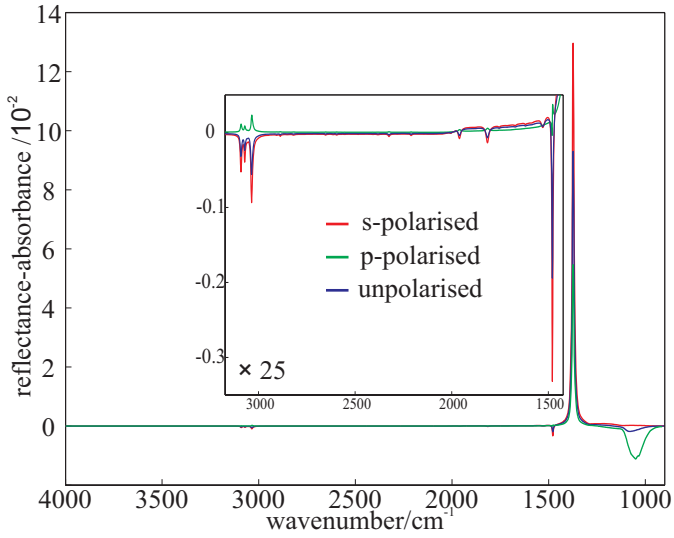


Figure 2.16: IRRA spectrum for benzene on glass at an incidence angle of 80°. The inset shows the region around 3000–1500 cm^{-1} expanded 25 times to show the benzene bands

Other interesting features relating to different polarisations have been alluded to earlier. A glass substrate permits significant s-polarised as well as p-polarised spectra. Since unpolarised radiation can be regarded as being composed of s- and p-polarised components in equal amounts, the unpolarised spectrum is the average of those obtained from the two linear polarisations. At 80°, s-polarised radiation gives positive analyte bands to the blue of the glass feature whereas p-polarised radiation gives negative ones (see the left-hand panel of Figure 2.17 for an expanded view of the C-H stretching bands near ~3000 cm^{-1}). The analyte bands in the unpolarised spectrum are therefore weakened by cancellation of the two polarised components, which has two major consequences: the first is that an unpolarised spectrum from a glass surface is weaker than that for the same analyte on a metal surface; the second is that the signal to noise ratio is slightly degraded.

It should be noted that the sign of the bands in the p-polarised spectrum are also dependent on the incidence angle and the substrate and analyte refractive indices. At incidence angles less than the Brewster angle, the sign of the p-polarised spectrum matches that for the s-polarised spectrum as shown in the right-hand panel of Figure 2.17.

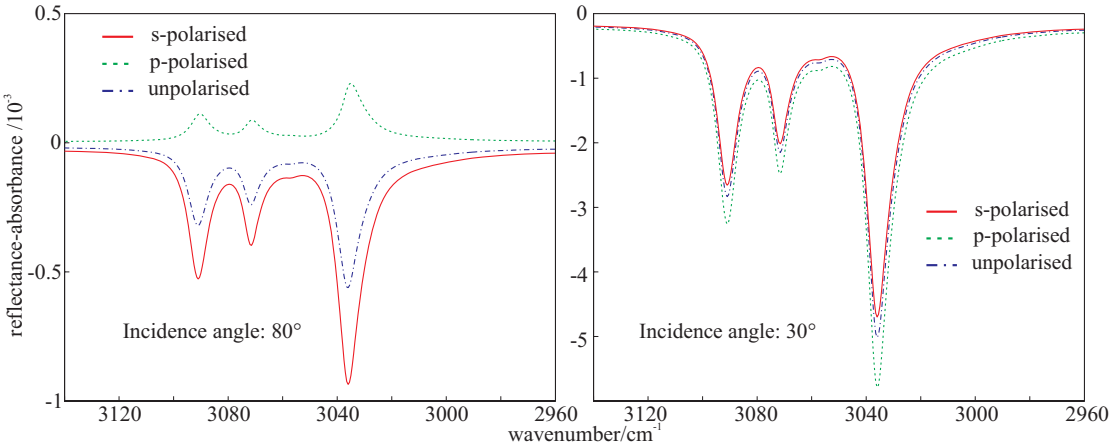


Figure 2.17: Enlarged section of the IRRA spectrum for a 10 nm film of benzene on glass showing the CH region for the incidence angles of 80° on the left and 30° on the right. These plots clearly show the difference in sign for the p-polarised spectrum and also the difference in signal intensity.

2.4.6 Order within the analyte film

The methods of film preparation used in the work presented in this thesis might not be expected to produce ordered films, but analysis of the results in Chapter six provides evidence for partial preferential orientation for sodium dodecyl sulfate on stainless steel. Consequently, a brief description of some of the affects of ordering is warranted here; further information can be found in cited references.[14,21,24] Anisotropic, conformationally ordered surface films can be produced on solid substrates by using techniques such as self assembly and Langmuir-Blodgett methods and many IRRAS studies have been conducted on films prepared in this way.[14, 21, 24, 25] The existence of such order can induce changes in the RA spectra beyond those discussed above, including band shape distortions.[14] For dielectric substrates, the extent of band distortions in IRRAS is very dependent on the polarisation and the incidence angle. It has been shown that distortions can be minimised for spectra collected at the Brewster angle,[25] although this does have the adverse effect of lowering the signal to noise ratio.[25] The optical effects that influence the band position, shape and intensity, have been shown to be very sensitive to both *n* and *k* as well as the wavelength.

2.5 Quantitative applications of IRRAS

Since quantitative IRRAS is the technique employed in the research covered in this thesis, a brief discuss of quantitative applications is presented here. The potential for quantitative applications of IRRA spectra has been suggested by theoretical calculations and demonstrated in a number of experimental studies.[9, 16, 24-26] The intensity of an IRRA spectrum is related to thickness of the analyte film. For films much thinner than the wavelength of the interrogating radiation, this quite linear, and simple Beers-law-type quantification can be used. However, as the film thickness approaches the radiation wavelength, interference phenomena become important and the linearity of the RA signal deteriorates.[9] A study by Allara and Swalen[24] investigated IRRAS of cadmium arachidate films, prepared by using Langmuir-Blodgett techniques, on a silver substrate. The linear range extended from a 1–11 monolayers, which (with a layer thickness determined to be 27 Å) corresponded to a maximum linear-range thickness of ~30 nm. An important point about this and many of the previous studies was that the systems were prepared in a very controlled fashion that ensures flat surfaces and consistent and precisely known surface coverages. It also provides an easy quantitative reference for how much analyte is present on the surface. Another commonly used preparation technique is self assembly as this also leads to the formation of homogeneous films.

The first studies reporting a quantitative relationship were on metallic substrates, which (as noted above) are simpler than dielectric systems. But over time, such studies have been extended to dielectric surfaces. In an example is presented by Niemczyk and coworkers,[25, 27] borophosphosilicate glass (BPSG) films were prepared on silicon wafers. IRRA spectra were analysed and were shown to display a linear relationship between the RA intensity and boron and phosphorous content in the film. Also discussed are the changes in quantitative nature with varying incidence angle. The IRRA technique is an attractive one because it is non destructive and spectra can be collected directly from a surface. Before IRRAS was introduced infrared transmission spectra were used which meant that films needed to be prepared on special wafers which where transparent to IR radiation.

2.6 Conclusions

The discussions in this chapter relate to ideal systems where the thin film covers the entire substrate surface with a uniform thickness and the air/film and film/substrate interfaces are parallel and optically flat. Most of the experimental literature referenced in this chapter relates to analyte films that were created using Langmuir-Blodgett and self-assembly techniques, which permit highly controlled film preparation. Such conditions are only approximated in situations relevant to pharmaceutical manufacturing equipment. Although it is relatively easy to obtain flat glass, this is a much greater problem for metals, while other substrates, such as silicone, are porous. It is to be expected that the degree of surface roughness will have a bearing on the homogeneity of surface coverage and structural order of surface species, which in turn could alter the potential for excitation of various vibrational modes. This will be discussed further in Chapter six.

Neither Langmuir-Blodgett nor self-assembly techniques were used to create the surface films in this research. To better model the situation present when an analyte dries on a surface such as a pharmaceutical vat, spraying and smear methods were utilized, as described in more detail in Chapter four. Images obtained using a scanning electron microscope (SEM) show that these methods cause “polka dot” patterning of the analyte on the substrate surface, rather than a uniform film. It is this aspect of the research which makes the quantitative nature of this project more uncertain. More complex theoretical treatments, such as effective medium theory (EMT) and the Bruggeman formula, can be used to model inhomogeneous films with islands of analyte on a support substrate.[21] The theory presented here is therefore only an approximation to the nature of the surface loadings created during this research although the preceding research has shown it to be adequate to describe the trends seen in the collected spectra.

2.7 References

1. Gauglitz, G. and Vo-Dinh, T. *Handbook of Spectroscopy*; Wiley-VCH: Weinheim, 2003.
2. Guyt, J. M., Van Eesbeek, M. and Van Papendrecht, G. Detection of organic contamination on surfaces by infrared spectroscopy, *Proceedings of SPIE - The International Society for Optical Engineering*, **2002**, *4774*, 262-271.
3. Griffiths, P. R. and de Haseth, J. A. *Fourier Transform Infrared Spectroscopy*; John Wiley & Sons: New York, 1986.
4. Smith, B. C. *Fundamentals of Fourier Transform Infrared Spectroscopy*; CRC Press: New York, 1996.
5. Eischens, R. P. and Pliskin, W. A. In *Advances in Catalysis*; Eley, D. D., Frankenburg, W. G. and Komarewsky, V. I.; Academic Press inc.: New York, 1958; Vol. X.
6. Greenler, R. G. Infrared study of adsorbed molecules on metal surfaces by reflection techniques, *Journal of Chemical Physics*, **1966**, *44*, 310-315.
7. Pickering, H. L. and Eckstrom, H. C. Heterogeneous reaction studies by infrared absorption, *Journal of Physical Chemistry*, **1959**, *63*, 512-517.
8. Francis, S. A. and Ellison, A. H. Infrared spectra of monolayers on metal mirrors, *Journal of the Optical Society of America*, **1959**, *49*, 131-139.
9. Greenler, R. G., Rahn, R. R. and Schwartz, J. P. Effect of index of refraction on position, shape, and intensity of infrared bands in reflection-absorption spectra, *Journal of Catalysis*, **1971**, *23*, 42-48.
10. Bertie, J. E. In *Handbook of Vibrational Spectroscopy: Theory and Instrumentation.*; Chalmers, J. M. and Griffith, P. R.; John Wiley and Sons, LTD: Chichester, UK, 2002; Vol. 1, 88-100.
11. Greenler, R. G. Reflection method for obtaining the infrared spectrum of a thin layer on a metal surface, *Journal of Chemical Physics*, **1969**, *50*, 1963-1968.
12. Greenler, R. G. Design of a reflection-absorption experiment for studying IR-spectrum of molecules adsorbed on a metal-surface, *Journal of Vacuum Science & Technology*, **1975**, *12*, 1410-1417.

13. Porter, M. D. IR external reflection spectroscopy: a probe for chemically modified surfaces, *Analytical Chemistry*, **1988**, *60*, 1143A-1155A.
14. Mielczarski, J. A. External reflection infrared spectroscopy at metallic, semiconductor, and nonmetallic substrates. 1. monolayer films, *Journal of Physical Chemistry*, **1993**, *97*, 2649-2663.
15. Blaudez, D., Buffeteau, T., Desbat, B., Fournier, P., Ritcey, A.-M. and Pezolet, M. Infrared reflection-absorption spectroscopy of thin organic films on nonmetallic substrates: optimal angle of incidence, *Journal of Physical Chemistry B*, **1998**, *102*, 99-105.
16. Umemura, J., Kamata, T., Kawai, T. and Takenaka, T. Quantitative evaluation of molecular orientation in thin Langmuir-Blodgett films by FT-IR transmission and reflection-absorption spectroscopy, *Journal of Physical Chemistry*, **1990**, *94*, 62-67.
17. Chollet, P. A., Messier, J. and Rosilio, C. Infrared determination of the orientation of molecules in stearamide monolayers, *Journal of Chemical Physics*, **1976**, *64*, 1042-1050.
18. Allara, D. L. A summary of critical issues for application of spectroscopy to characterisation of surface processing, *Critical Reviews in Surface Chemistry*, **1993**, *2*, 91-110.
19. Allara, D. L., Baca, A. and Pryde, C. A. Distortion of band shapes in external reflection infrared spectra of thin polymer films on metal substrates, *Macromolecules*, **1978**, *11*, 1215-1220.
20. Heavens, O. S. *Optical Properties of Thin Solid Films*; Dover Publications, Inc.: New York, 1991.
21. Tolstoy, V. P., Chernyshova, I. V. and Skryshevsky, V. A. *Handbook of Infrared Spectroscopy of Ultrathin Films*; Wiley-Interscience: Hoboken, New Jersey, 2003.
22. Graham Smith, F. and King, T. A. *Optics and Photonics: An Introduction*; Wiley & Sons, Ltd: Chichester, UK, 2000.
23. Perston, B. B. Ph.D. Thesis, University of Canterbury, Christchurch, 2006.
24. Allara, D. L. and Swalen, J. D. An infrared reflection spectroscopy study of oriented cadmium arachidate monolayer films on evaporated silver, *Journal of Physical Chemistry*, **1982**, *86*, 2700-2704.

25. Niemczyk, T. M., Zhang, S. B., Franke, J. E. and Haaland, D. M. Quantitative determination of borophosphosilicate glass thin-film properties using infrared emission spectroscopy, *Applied Spectroscopy*, **1999**, *53*, 822-828.

26. Gaillard, F., Linossier, I., Sweeney, M., Reffner, J. A. and Romand, M. Grazing angle micro-FTIR spectroscopy (GAM-FTIR): applications to adhesion studies, *Surface and Interface Analysis*, **1999**, *27*, 865-870.

27. Zhang, L., Franke, J. E., Niemczyk, T. M. and Haaland, D. M. Optimized external IR reflection spectroscopy for quantitative determination of borophosphosilicate glass parameters, *Applied Spectroscopy*, **1997**, *51*, 259-264.

Chapter 3

Chemometrics

3.1 What is chemometrics?

Chemometrics has been described by Svante Wold as the art of extracting chemically relevant information from data produced in chemical experiments.[1] This is analogous to other fields, which have similar topics such as biometrics and econometrics. Chemometrics uses statistics, along with numerical and operational analysis, to build mathematical models for experimental data. These models use multivariate (rather than univariate) analysis to describe complicated systems subject to the effects of multiple variables. The possible statistical methods are numerous and include classical least squares (CLS), inverse least squares (ILS), principal component regression (PCR) and partial least squares (PLS).

3.2 An introduction to multivariate analysis

Traditional univariate methods are commonly used for spectroscopic analysis. They involve the measurement of a single spectral aspect, such as the peak height or area of a band. The measurements are then related (usually) to component concentration using a simple relationship such as Beer's law.

In order for univariate methods to be effective, the analysed spectra must conform to certain criteria. The relevant band of the analyte must be isolated (or nearly isolated) from other spectral features so that height or area measurements can be performed without interferences. The analysed band needs to be free of baseline variations, which can detrimentally affect reproducibility and the reliability of the results obtained. There are also other considerations such as non-linearities arising from the instrument or chemical

interactions.[2] Unfortunately, collecting spectra that conform to these criteria is often difficult. In some cases, spectra can be manipulated using various methods that improve reproducibility and linearity. Such manipulations are commonly associated with baselines and involve the fitting of linear or polynomial functions. Even using correction or isolation methods it is still difficult to obtain IR spectra that can be used for quantitative univariate analysis. That, in turn, has meant there has generally been a preference for using other analytical methods, such as chromatography and UV/visible spectroscopy. The introduction of multivariate methods to chemical analysis has meant that the positive qualities of FTIR spectra can be utilised even when spectra include overlapping bands and baseline variations.

Multivariate spectroscopic analysis differs from univariate analysis in that information from two or more frequencies are investigated simultaneously, improving both sensitivity and analyte selectivity. Multivariate analysis usually assumes a linear relationship, such as Beer's law. Consider the case where spectra are measured from m samples, each containing l distinct components, over n wavelengths. The spectra can be assembled into an $m \times n$ absorbance matrix, **A**, whose rows relate to each sample and columns to each wavelength.[*] If **K** is an $l \times n$ matrix of the products of the path lengths and absorbance coefficients for each component at each wavelength and **C** is an $m \times l$ concentration matrix, then Beer's law gives equation (3.1), (**K** can also be thought of as the pure component spectra at unit concentration for the relevant path length).

$$\mathbf{A} = \mathbf{C}\,\mathbf{K} \tag{3.1}$$

Equation (3.1) is the ideal equation for the case where all components within a system are known and behave ideally, and the spectra are collected at infinite resolution with no errors.

[*] The mathematical conventions used throughout this report are that bold capital letters represent matrices, bold lower case letters represent vectors and italics represent scalars. Spectra are also expressed as row vectors; this decision is arbitrary and both column and row treatments are used in the literature.

In order to describe experimentally collected spectra which contain unknowns, baseline effects and instrument errors, equation (3.1) can be modified to include an $m \times n$ error matrix $\mathbf{E_A}$, as shown in equation (3.2).

$$\mathbf{A} = \mathbf{C}\,\mathbf{K} + \mathbf{E_A} \tag{3.2}$$

A multivariate analysis method based on equation (3.2) would attempt to determine the concentration matrix from the spectra. Essentially, a calibration set of samples with known **A** and **C** matrices would be used to determine a model **K** matrix, which would then be used to determine the unknown concentrations from measured spectra. By analysing more than one wavelength at a time, interfering components could be detected, and baseline variation compensation could be achieved.

The first multivariate methods to be developed were inverse least squares (ILS) and classical least squares (CLS). These use multiple linear regression (MLR) to analyse spectra using a number of different wavelengths simultaneously. The following sections describe how CLS and ILS methods work and how they can be developed, using factor analysis, into the more robust principal component regression (PCR) and partial least squares (PLS) methods.[†]

3.3 Data pre-treatments

Before spectra or concentration data are subjected to the calibration process they can be pre-treated. Some pre-treatments are intended to reduce or remove variations between spectra that could lead to errors when the data are subjected to the calibration method; others are designed to improve the linearity between spectra and concentration. Some pre-treatments are so standard that they are hardly recognised as such. One example is the

[†] The treatment used is similar to that which is presented in 3. Franke, J. E. In *Handbook of Vibrational Spectroscopy: Sample Characterization and Spectral Data Processing*; Chalmers, J. M. and Griffith, P. R.; John Wiley and Sons, LTD: Chichester, UK, 2002; Vol. 3, 2276-2292.

conversion from spectral transmittance to absorbance, which is vastly more linear with concentration.

It is important that any pre-treatment is appropriate and applied correctly to preclude degradation in the relationship between the spectra and the concentrations. If a pre-treatment can be performed in such a way that it cleanly removes an artefact, then it is usually considered good practice to perform it, rather than to rely on the model to fit it. This is especially true for non-linear artefacts that the model may not be able to fit.[4]

Pre-treatments can generally be classified into three categories of transformations: local, which work on single values; neighbourhood based, which work over a spectral window (examples include smoothing or derivatives); and global, where all variables are transformed.[5] Depending on which techniques are employed, different types of variation can be removed.

Smoothing is a technique that is used to remove high-frequency noise. Simple smoothing involves averaging several consecutive readings around a central wavelength. The window is moved along the spectrum until all wavelengths have been treated. This technique works well, but has the disadvantage that it broadens the spectral bands. To reduce this effect, weighted averages can be taken where points closest to the central wavelength contribute more to the average than those further away.

Spectra can be scaled for situations where different regions have greater or lesser influence in the calibration. The simplest application of this type is where the regions of a spectrum are multiplied by 1 or 0 depending on whether they are to be included in the calibration or not. This allows selection of regions where there is a good relationship between concentration and spectral features and removal of regions that contain interfering components.

Variance scaling is a technique that allows one to equalise the influence of each analyte in the calibration. It involves adjustment of variables, such as spectral intensity or

concentration, so that the distributions for each analyte have equal variance. Variance scaling does come with some warnings. If the technique is applied in a poor situation (for example where there are analytes whose spectral intensity is very different) it is likely to introduce an undesirable sensitivity to changing conditions,[4] since variables with intense signals have their influence reduced while variables with weak signals, which contain mostly noise, have their influence increased. The technique is applied on a variable-by-variable basis, and can be applied to concentration or spectra data matrices, either separately or together.

Mean centring is another commonly used technique that can be applied to the absorbance and/or concentration matrices. It is applied variable-by-variable; in other words wavelength by wavelength for the spectral data, or analyte by analyte for the concentration data. For the spectral data, the average absorbance at each wavelength is calculated over all spectra. This average is then subtracted from all spectra. The same process can be repeated for the concentration data, where the average concentration for a particular analyte is subtracted from all the concentrations of that analyte. Mean centring moves the origin of the coordinate system to the centre of the data set in an attempt to prevent data points that are further away from the origin from exerting undue leverage (influence on the calibration). The leverage of a data point is directly proportional to its distance from the origin in the factor space. In the absence of mean centring, high concentration values have the greatest leverage. But with mean centring, data points at low and high concentration have similar leverages. Mean centring removes a degree of freedom from the data matrices, allowing the calibration to focus more on differences between the data points. However, it also represents loss of information about the relative magnitude of the eigenvalues and relative errors.[4] The data are mapped onto an abstract space where the relative positions of data points are unchanged, but the origin has no external physical or chemical significance.[4] This pre-treatment is very common and some computer packages perform it automatically. This does not mean that it is appropriate for all data sets and, like all pre-treatments, the properties of the data should be considered before its application.

Other pre-treatments, such as derivatisation and normalisation, are used to remove low frequency noise. Derivatives are usually applied to spectra to correct for baseline variations. First-order derivatives correct for constant offsets (similar to subtraction of the mean absorbance from each spectrum), while higher-order derivatives are used to correct for more complicated baselines; second derivatives for linear baselines; third-order derivatives for quadratic baselines, and so forth.[6] Derivative calculations have no effect on differences due to sensitivity changes, but they do emphasize high-frequency noise, therefore smoothing is also usually applied to the transformed spectra.[6]

The definition of normalisation can vary with situation. In this thesis, it means that the data set has been adjusted so that each spectrum in the set has the same intensity. Normalisation is more commonly used in qualitative analysis since intensity information about the spectral components between spectra is lost.

It is very important to realise that pre-treatments applied to the calibration data matrices must also be applied to unknown samples.

3.4 Classical least squares

CLS assumes that the absorbance of a sample component is proportional to the concentration of that component, and that the overall spectrum is therefore the concentration-weighted sum of the spectra of all components. Since the spectra will also contain errors, equation (3.2) pertains to this case. The aim of the CLS analysis is to calculate the **K** matrix, where the sum of squares of the spectral residuals is minimised, and so it is for this reason that CLS is sometimes referred to as **K**-matrix calibration. The sum of squared residuals is given by $\mathbf{E_A}^T\mathbf{E_A}$;‡ an expression for which is shown in equation (3.3), which starts from the relationship in equation (3.2) and assumes that the error in $\mathbf{E_A}$ is random (with expectation value of zero) so that $\mathbf{A}^T\mathbf{CK}$ is symmetric. It is

‡ The superscript T refers to transpose operation where the rows and columns of the given matrix are interchanged. The transpose operation in other texts is sometimes denoted by ' rather than ^{T}. 7. Draper, N. R. and Smith, H. *Applied Regression Analysis*, 2nd ed.; John Wiley and Sons, Inc: New York, 1981.

also assumed that residual spectral errors in $\mathbf{E_A}$ are: i) normally distributed around a mean of zero; ii) independent; and iii) exhibit equivalent variances (homoscedasticity).

$$\begin{aligned} \mathbf{E_A^T E_A} &= \mathbf{(A - CK)^T (A - CK)} \\ &= \mathbf{A^T A - A^T CK - K^T C^T A + K^T C^T CK} \\ &= \mathbf{A^T A - 2A^T CK + K^T C^T CK} \end{aligned} \quad (3.3)$$

Using partial derivatives, equation (3.3) can be solved for **K** while minimising $\mathbf{E_A^T E_A}$. This is achieved by setting $\mathbf{E_A^T E_A} = 0$ and differentiating with respect to **K** (a more complete description can be found in references[3, 7]). This treatment leads to equation (3.4), which is known as the normal equation and represents a set of simultaneous linear equations in **A** and **C** that yield the least squares solution for **K**. In other words $\mathbf{E_A}$ can be eliminated from equation (3.2), since it is assumed to be zero, and the system can be solved to give the least squares solution for **K** (equation (3.5)).

The **K** matrix calculated in equation (3.5) is a least squares estimation of the pure component so that the sum of squared spectral errors over all elements of **A** is minimised. The CLS method achieves this by building up the **K** matrix one wavelength at a time for all concentration data for all samples.

$$\mathbf{C^T A} = \mathbf{C^T CK} \quad (3.4)$$

$$[\mathbf{C^T C}]^{-1}\mathbf{C^T A} = [\mathbf{C^T C}]^{-1}\mathbf{C^T CK}$$

$$[\mathbf{C^T C}]^{-1}\mathbf{C^T A} = \mathbf{K} \quad (3.5)$$

The above series of equations can only be solved if $[\mathbf{C^T C}]^{-1}\mathbf{C^T}$, the pseudo inverse of **C**,[§] actually exists. For this to be true, the determinant of $[\mathbf{C^T C}]^{-1}$ must be non-zero and therefore **C** must have at least as many rows as columns. In practical terms this means that the calibration set has to contain at least as many samples as there are components. This seems to be a reasonable, and possibly self-evident, constraint, but it should be

[§] The pseudo inverse is also known as the Moore-Penrose generalised pseudo inverse. 2. Danzer, K., Otto, M. and Currie, L. A. Guidelines for calibration in analytical chemistry part 2. Multispecies calibration (IUPAC Technical Report), *Pure and Applied Chemistry*, **2004**, *76*, 1215-1225.

noted that the number of components does not just refer to chemical species but can also include baseline and instrument effects. Another important condition for the pseudo inverse of **C** to exist is that the samples need to be mutually independent. If there is any linear dependence (e.g. if all concentrations in two samples are related by the same ratio) then $[\mathbf{C}^T\mathbf{C}]$ will be singular and its inverse will not exist. If some rows of **C** are almost dependent, then the pseudo inverse of **C** can be calculated, but the solution may be unstable and give rise to poor results. There are statistical indicators, such as condition numbers, that can be used to assess the stability of an inverse.[8]

Once **K** has been calculated it can be used to predict the concentration of components in unknown samples (whose matrices are indicated by the subscript "unk").

$$\mathbf{A}_{unk} = \mathbf{C}_{unk}\mathbf{K} \tag{3.6}$$

$$\mathbf{A}_{unk}\mathbf{K}^T = \mathbf{C}_{unk}\mathbf{K}\mathbf{K}^T$$

$$\mathbf{A}_{unk}\,\mathbf{K}^T[\,\mathbf{K}\mathbf{K}^T]^{-1} = \mathbf{C}_{unk}\mathbf{K}\mathbf{K}^T[\mathbf{K}\mathbf{K}^T]^{-1}$$

$$\mathbf{A}_{unk}\,\mathbf{K}^T[\,\mathbf{K}\mathbf{K}^T]^{-1} = \mathbf{C}_{unk} \tag{3.7}$$

Since this derivation is very similar to the derivation for **K** in equation (3.5), there are similar constraints on **K** as there were on **C**. For the pseudo inverse of **K** to exist there must be at least as many wavelengths as there are components. Again, this is reasonable, and most spectra have a large number of wavelengths compared to the number of components.

Equation (3.7) can be simplified in terms of the $\mathbf{K}_{cal}$ matrix (equation (3.8)), also known as the calibration or regression matrix, to give equation (3.9).

$$\mathbf{K}_{cal} = \mathbf{K}^T[\mathbf{K}\mathbf{K}^T]^{-1} \tag{3.8}$$

$$\mathbf{C}_{unk} = \mathbf{A}_{unk}\mathbf{K}_{cal} \tag{3.9}$$

An **E** matrix is not present in equation (3.6) since spectral errors are minimised. However, a residual matrix may be calculated (equation (3.10)) and used for outlier

detection and to determine how well the calibration fits the unknown spectra: if $\mathbf{A}_{\mathbf{unk}}$ is the matrix of unknown spectra, then $\hat{\mathbf{A}}_{\mathbf{unk}}$, the CLS estimate of the unknown spectra, can be determined according to equation (3.11).

$$\mathbf{E}_{\mathbf{unk}} = \mathbf{A}_{\mathbf{unk}} - \hat{\mathbf{A}}_{\mathbf{unk}} \tag{3.10}$$

$$\hat{\mathbf{A}}_{\mathbf{unk}} = \mathbf{A}_{\mathbf{unk}} \mathbf{K}_{\mathbf{cal}}\mathbf{K} \tag{3.11}$$

CLS models are suitable for systems where all the components are known, their appearance in the spectra is predictable and usually linear, and any errors in the concentration values are small compared to those in the spectra. To construct a system that conforms to these requirements can be a difficult task. For example, many samples will contain other "components" whose effects on the spectra are more difficult to quantify. These can include baseline effects, temperature changes, instrument drift and operator effects. Sometimes their effects can be compensated for by including extra terms in equation (3.2), as demonstrated in equation (3.12) where $\mathbf{A}_0$ and $\mathbf{C}_0$ are mean-centred absorbance and concentration matrices, respectively. For example, in the case of variable baseline offsets, $\mathbf{G} = \mathbf{g}$ would be a column vector of intercept coefficients for each spectrum, and $\mathbf{L} = \mathbf{l}$, a row vector of ones that represent the shape of the offsets. This could be extended to $\mathbf{G}$ being an $m \times 2$ matrix of intercept and slope coefficients and $\mathbf{L}$ being a $2 \times n$ matrix of intercept and slope etc.

$$\mathbf{A}_0 = \mathbf{GL} + \mathbf{C}_0\mathbf{K} + \mathbf{E}_\mathbf{A} \tag{3.12}$$

CLS does have a number of positive features; it uses the full spectral range, unlike univariate methods and some other multiple linear regression methods such as ILS (explained in the next section). The spectral residuals ($\mathbf{E}_\mathbf{A}$) and the estimated pure component spectra in the $\mathbf{K}$ matrix can be used to yield information for outlier detection and to provide qualitative information such as: the presence of molecular interactions; spectrometer non-linearities; the presence and sometimes the identification of unexpected components; components whose concentrations are changing; the identification of reaction products; and information which can aid in the assignments of spectral bands.[9]

The major disadvantage with CLS is the requirement for a well-defined system with detailed knowledge of either the concentrations or pure component spectra for all chemical species, along with instrumental, environmental and physical factors that will influence the spectral absorbance. In addition, in most CLS models, the relationships between absorbance and concentration are assumed to be strictly linear.

3.5 Inverse least squares

ILS gained its name though the use of the inverse Beer's law relationship expressed in equation (3.13). **C** and **A** are as above, **Φ** is an $n \times l$ matrix that relates the component concentrations to the spectral intensity at each wavelength (the more commonly used symbol, **P**, is used later in this chapter to designate the loadings matrix in PLS) and $\mathbf{E_C}$ is an $m \times l$ matrix of concentration residuals. ILS estimates a **Φ**-matrix to give a product **AΦ** that is as close as possible to the original **C** in the least squares sense; hence ILS is also known as **Φ**-matrix calibration.

$$\mathbf{C} = \mathbf{A\Phi} + \mathbf{E_C} \tag{3.13}$$

ILS is similar to CLS, but it operates on the minimisation of concentration residuals rather than spectral residuals. The similarity means that equation derivations are analogous and the assumptions made about the spectral residuals in CLS now applying to the concentration residuals. The resultant normal equations for the ILS are shown in equation (3.14) and can be solved to give the least squares solution for **Φ** (equation (3.15)).

$$\mathbf{A^T C} = \mathbf{A^T A\Phi} \tag{3.14}$$

$$[\mathbf{A^T A}]^{-1}\mathbf{A^T C} = [\mathbf{A^T A}]^{-1}\mathbf{A^T A\Phi}$$

$$[\mathbf{A^T A}]^{-1}\mathbf{A^T C} = \mathbf{\Phi} \tag{3.15}$$

Since the pseudo inverse of **A** must be determined, **A** is subject to analogous constraints to **C** in CLS: there must be at least as many samples as wavelengths and at least as many wavelengths as components. The second constraint is not difficult to accommodate, but

the first of these leads to a major problem for ILS since it is unlikely that the number of samples will be greater than the number of wavelengths. There are several ways around this problem.[6, 9] The first is to select a number of individual wavelengths that is less than the number of samples collected. Optimised wavelength selection is important, as poor selections can lead to the overfitting of the spectral data and/or poor baseline fitting, which can lead to a near-singular **A**-matrix with an unstable inverse. In some situations, inclusion of too many wavelengths can cause the model precision to be degraded by collinearity problems.[9] The second method is to condense the data so that there are fewer wavelength measurements than samples. Both methods lead to the major disadvantage of ILS – that (unlike CLS) it is not a full-spectrum method and does not use all the spectral information that is available.

ILS calibration works in a similar way to CLS, but instead of estimating pure component spectra, concentrations are fitted one analyte at a time to the spectral intensities at selected wavelengths over all samples. This results in an estimation of the regression matrix $\mathbf{\Phi}$ in which each column relates to one analyte. This fit is performed in such a way that the total sum of squared concentration errors $\mathbf{E_C}^T\mathbf{E_C}$ is minimised.

Once $\mathbf{\Phi}$ has been calculated, it can be applied to unknown spectra and used to predict component concentration, $\hat{\mathbf{C}}_{\mathbf{unk}}$.

$$\hat{\mathbf{C}}_{\mathbf{unk}} = \mathbf{A}_{\mathbf{unk}}\mathbf{\Phi} \tag{3.16}$$

Substituting $\mathbf{\Phi}$ in the inverse Beer's law (equation (3.13)) gives rise to equation (3.17), with the additional incorporation of error matrix, $\mathbf{E_C}$.

$$\hat{\mathbf{C}}_{\mathbf{unk}} = \mathbf{A}_{\mathbf{unk}}\mathbf{\Phi} + \mathbf{E_C} \tag{3.17}$$

One of the major advantages of ILS over CLS is that not all absorbances need to be allocated to calibrated components; only those analytes of interest are required to have known concentrations in the calibration. This is made possible by assuming that the analyte concentrations in **C** are independent, so **C** can be reduced to a column vector, **c**,

that contains the concentration information for a single analyte. Equation (3.15) then results in a $\boldsymbol{\varphi}$ ($n \times 1$) vector, which can be used for the prediction of a single analyte in an unknown spectrum. The requirement that unknown spectra should be similar to the calibration spectrum is still important. In other words, the unknown samples should not contain new components that were not present in the calibration samples and the concentrations of these components should not be outside the range of the calibration samples.

In ILS, errors are assumed to be in the concentration matrix rather than in the absorbance matrix (as in CLS). This conveys an advantage in that not all components need to be calibrated. If there are regions of nonlinearity in the spectra, the model may be able to utilise other regions of the spectrum where there is a good correlation between absorbance and concentration while, at the same time, effectively ignoring other components.

ILS can be extended to accommodate some deviations from the initial assumptions by using weighted least squares. It is common for concentration values to be heteroscedastic (non-homoscedastic); for example, quite typically as the concentration increases its absolute error is likely to increase. Equation (3.15) can be modified to include an $m \times m$ diagonal matrix, $\mathbf{V}$, containing concentration variance errors for the m calibration samples, as shown in equation (3.18). $\mathbf{V}^{-1}$ (the diagonal matrix with elements V_{jj}^{-1} for $j = 1, 2, 3 \ldots m$) scales the data relative to the uncertainties in the concentration. This reduces to the ordinary least squares solution seen earlier (equation (3.15)) in the limit that $\mathbf{V}$ tends to the identity matrix.

$$[\mathbf{A}^{\mathbf{T}}\mathbf{V}^{-1}\mathbf{A}]^{-1}\mathbf{A}^{\mathbf{T}}\mathbf{V}^{-1}\mathbf{C} = \boldsymbol{\Phi}_{\mathbf{WLS}} \qquad (3.18)$$

If there are concentration errors in $\mathbf{E}_{\mathbf{C}}$ that are correlated between some samples or for various wavelengths, then a generalised least squares method can be used to weight the samples. $\mathbf{V}$ is replaced with $\boldsymbol{\Sigma}$, which has the same dimensions but now contains non-zero off-diagonal elements. Similar treatments can be performed in CLS calibrations for non-homoscedastic effects related to wavelengths.

Because ILS models do not use the full spectral range, they tend to perform best with strong spectra and minimal band overlap. The model can tend to overfitting, which can reduce precision and robustness. In an attempt to avoid overfitting, a large number of samples is used to minimise the undue influence by spurious individual sample features above the overall trends present in all samples. ILS calibrations only yield information on the concentration of the analyte or the physical property for which the model was calibrated and the concentration residuals do not give any information on outliers or the shape of the analyte spectrum. ILS calibration can be seen as wasteful since not all of the spectral information contained in the spectrum is used.

3.6 Factor analysis

As discussed above, both CLS and ILS have advantages and disadvantages. Methods like principal component regression (PCR) and partial least squares (PLS), which are discussed in the next sections, attempt to avoid the disadvantages while combining the advantages. A major advantage of CLS is that it is a full spectrum method. ILS is not a full spectrum method but, unlike CLS, it does not require information on all components that contribute to the spectra. To combine these positive features, factor analysis (also known as principal component analysis or PCA) methods are used.

To understand how factor analysis works, consider the synthetic IR spectra (126 wavelength data) for two analytes shown in Figure 3.1. These spectra contain overlapping bands that would make univariate analysis methods problematic. Twenty spectra for mixtures of the two components were calculated from these pure analyte spectra using equation (3.1), where **K** is a (2 × 126) matrix of absorbances of the pure analytes and **C** is a randomly generated (20 × 2) matrix of concentration values for each analyte.[**]

[**] Noise could also have been added to the **A-** and **C**-matrices to simulate errors in the spectrometer or the primary calibration method, but this was not carried out here.

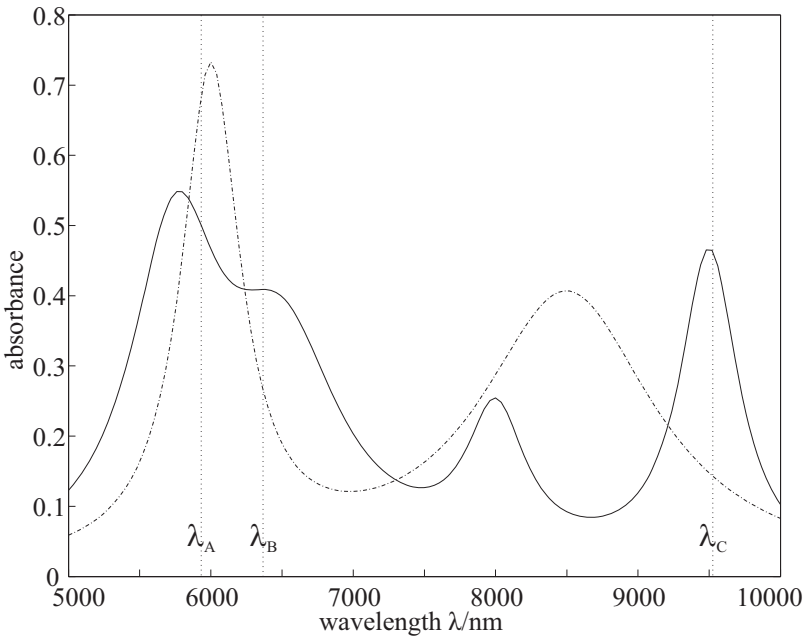


Figure 3.1: Synthetic IR spectra for analytes 1 and 2 with overlapping bands. Wavelengths A, B and C were used to demonstrate how factor analysis works.

The absorbances for each spectrum at three wavelengths, λ_A, λ_B and λ_C (Figure 3.1), can be determined and used to generate a three dimensional plot (Figure 3.1), which shows that, for a system with only two components, all of the spectral vectors lie in a two-dimensional plane. If more wavelengths had been included, all points would still lie in a 2D-plane in a higher dimensional space. More generally, for *l* components whose spectra are measured over *n* wavelengths, the points will lie in an *l*-dimensional subspace of the *n*-dimensional space.

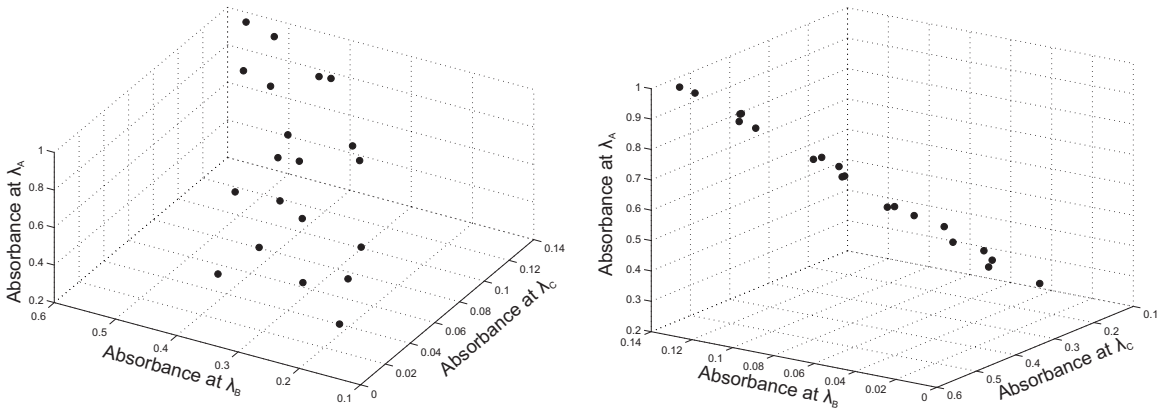


Figure 3.2: The absorbances at wavelengths λ_A, λ_B and λ_C of the mixture spectra containing analytes 1 and 2. The two plots show the same data from different angles making it easier to see that the spectral absorbances lie in a plane.

Only *l* vectors are required to completely describe an *l*-dimensional space if those vectors are chosen to be independent; such vectors can be used as axes for the subspace (they span the data space) and are known as a basis set. In chemometrics, they are known as 'factors' of the data space. Sometimes they are referred to as 'abstract factors' because they are not usually related to a single-component spectra; instead they represent linear combinations of the single-component spectra. By considering just the subspace, the dimensionality of the problem is reduced without loss of information. In the above example, the absorption spectrum of a particular mixture could be defined by two factors and the projections along each basis vectors rather than requiring 126 absorbance values.

The most common approach for constructing a basis set is to determine eigenvectors, which will be mutually orthogonal and therefore independent. The first eigenvector (or principal component) is the one which spans the greatest variance in the data; in other words it is the vector for which the sum of squares of the perpendicular distances to all data points is minimised. The second eigenvector is calculated to be orthogonal to the first and to span the most of the remaining variance, and so on. The associated eigenvalues provide information on the amount of variance spanned by the relevant eigenvector (generally defined as the sum of squares of the projections of the data onto the eigenvector) and they decrease as the order of the eigenvector increases. For the synthetic data shown in Figure 3.2, the first two eigenvectors would lie in the plane of the data and have non-zero eigenvalues. In the absence of noise, the third and successive

eigenvalues would be equal to zero. In the presence of weak noise, the later eigenvectors would be non-zero but small.

The last of these considerations points to a very elegant method for noise reduction. A spectrum is separated into parts according to its projection along the basis vectors (factors). The parts associated with factors with small eigenvalues are removed and the spectrum is reconstructed from those that remain. This process cannot remove all of the error in the data since some of it will be in the direction of the factors that were not removed. This is described in equation (3.19), where RE is the total or real error, IE is the imbedded error spanned by the retained factors, and XE is the extracted error (known as the residuals) that is removed by using a reduced number of factors.[4]

$$RE^2 = IE^2 + XE^2 \qquad (3.19)$$

The retained factors go under several other names, including basis vectors, principal components, loadings, latent variables and primary eigenvectors. The number of retained factors is called the rank or dimensionality of the data set. Ideally, the rank is equal to the number of components present in the spectra, which can include analyte species, chemical interactions, baseline variations and nonlinearities.

In a way, CLS uses an explicitly modelled factor analysis where the factors are constrained to obey Beer's law. Methods like PLS and PCR involve implicit modelling; since the factors are not constrained, they are abstract and lack some of the qualitative information present in CLS. Working in factor space with abstract factors does have advantages, perhaps the most important of which is the reduction of the dimensionality and therefore the noise, without comprising the information content. In addition, subjective assumptions about collinearity, linear independence of the spectra and the appropriate number of components can be avoided.

3.7 Principal component regression

PCR is a method for extracting the full potential from ILS calibrations which do not use all the spectral information. This is achieved by using PCA to reduce the dimensionality of the data set so that $r < m << n$ where r is the rank of the regenerated data set (remembering that m is the number of samples and n is the number of wavelengths).

For factor analysis to be performed on a matrix, it has to be square. The absorption matrix, **A**, has dimensions $m \times n$ where there are likely to be many more wavelengths than samples. Thus a covariant matrix is calculated, either as $\mathbf{A}\mathbf{A}^T$ ($m \times m$) or $\mathbf{A}^T\mathbf{A}$ ($n \times n$). Usually, the $m \times m$ matrix is used since no extra information is held in the larger $n \times n$ matrix. The eigenvectors (or loadings) are found (usually by using either singular value decomposition (SVD) or nonlinear iterative partial least squares (NIPALS) algorithms)[10] and collected in an $m \times n$ loadings matrix **R**, which has the same dimensions as **A**.

The next step is to discard the less meaningful (or "noise") factors, which are determined by assessing the size of the eigenvalues.[††] Once the number of significant factors has been decided, **R** can be reduced to $\mathbf{R}^+$ ($r \times n$). The original spectra in **A** are projected on to the new basis factors defined in $\mathbf{R}^+$, giving a score matrix, $\mathbf{A}_{\mathbf{proj}}$, with a reduced number of columns (equation (3.20)). $\mathbf{A}_{\mathbf{proj}}$ should contain all the significant spectral information that was present in **A** while also removing some spectral errors by reducing the number of eigenvectors.

$$\mathbf{A}_{\mathbf{proj}} = \mathbf{A}(\mathbf{R}^+)^T \tag{3.20}$$

$\mathbf{A}_{\mathbf{proj}}$ can be used to model **C** in a manner that is similar to that of an ILS calibration. This is shown in equation (3.21), where **D** is an $r \times l$ regression matrix, very similar to $\mathbf{\Phi}$ in ILS.

[††] Finding the boundary between useful and noise factors is a very important step in PCR and other calibration methods, such as PLS discussed in Section 3.9. The selection of the number of significant factors is discussed in Section 3.8 .

$$\mathbf{C} = \mathbf{A}_{\mathbf{proj}}\mathbf{D} \tag{3.21}$$

The reduced size of $\mathbf{A}_{\mathbf{proj}}$ ($r < n$) means it is possible to calculate the pseudo inverse of $\mathbf{A}_{\mathbf{proj}}$ and hence to solve for $\mathbf{D}$ (equation (3.22)).

$$[\mathbf{A}_{\mathbf{proj}}^{\mathbf{T}}\mathbf{A}_{\mathbf{proj}}]^{-1}\mathbf{A}_{\mathbf{proj}}^{\mathbf{T}}\,\mathbf{C} = \mathbf{D} \tag{3.22}$$

Before the unknown spectra are submitted to the calibration, a matrix $\mathbf{D}_{\mathbf{cal}}$ ($n \times l$) can be calculated.

$$\mathbf{D}_{\mathbf{cal}} = (\mathbf{R}^{+})^{\mathbf{T}}\mathbf{D} \tag{3.23}$$

The unknown spectra are collected in a matrix $\mathbf{A}_{\mathbf{unk}}$ (remembering that spectra are presented in rows). Multiplication of $\mathbf{A}_{\mathbf{unk}}$ by $\mathbf{D}_{\mathbf{cal}}$ (equation (3.24)) projects the spectra onto the factor space in $\mathbf{R}^{+}$, hence predicting the analyte concentration.

$$\hat{\mathbf{C}}_{\mathbf{unk}} = \mathbf{A}_{\mathbf{unk}}\mathbf{D}_{\mathbf{cal}} \tag{3.24}$$

PCR has obvious advantages over both CLS and ILS. It is a full spectrum method that does not require concentration information about all components that influence the spectrum. However, it does require that the unknown spectra do not contain influential but unknown components that were not present in the calibration spectra. A disadvantage of PCR is that it only uses the absorbance data in the **A**-matrix to determine the basis factors, and therefore noise removal is performed only on the **A**-matrix. It is therefore assumed that the information in the concentration matrix **C** is both accurate and free from errors. This is unlikely as the **C**-matrix also contain errors that are likely to be independent of the errors found in **A**. Another possible limitation of PCR is that, if the absorbance data are very collinear, an unstable calibration is possible. This can be best illustrated if a two component system is considered where the basis factors map out a plane. If the absorbance spectra are collinear then the angle between the two vectors is very small, which makes fitting a plane through the two vectors more susceptible to errors thereby producing a poor calibration.

3.8 Selection of the number of significant model factors

In both PCR and PLS, the selection of the rank (the number of significant factors) is very important. If too many factors are included, the model will be overfitted, meaning that some spectra with distinctive features may have an unreasonably large influence on the calibration. If insufficient factors are included, spectral information about the analytes will be lost and the resulting model will be inferior. Generally, the optimal rank depends on the number of independently varying chemical species and the number of other sources of systematic spectral variations.[11]

There are a number of ways of determining the optimal rank, and the choice of method requires an understanding of the types of errors present in the data. Section 3.6 introduced the concepts of total, imbedded and extractable error (equation (3.19)), which can be extended to relationships between different absorbance data.[10] Figure 3.3 is a mnemonic, developed by Malinowski, to represent these relationships.[12] Ideally one would like to determine RE (the real error) and remove it completely from the "raw spectra" to give the "pure spectra". The extractable error, XE, can be removed by reducing the number of factors and regenerating the spectra; but the remaining, imbedded error, IE, cannot be removed by this method without the loss of analyte information.

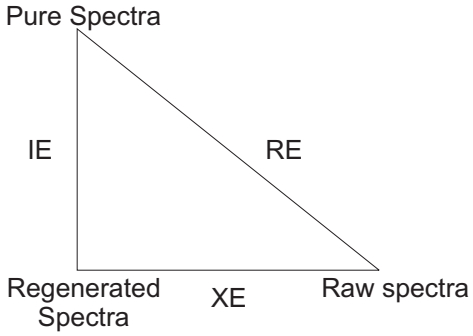


Figure 3.3: Mnemonic diagram developed by Malinowski to represent the error relationships present between three absorbance matrices.[12]

There are a number of methods that have been developed in attempts to calculate the RE, XE and IE, which are then used to estimate the optimal rank. These methods all assume

that errors are random and spherically distributed in the hyperspace of the spectra. Estimates of RE and XE require prior knowledge of the error levels in the spectra.[13] At the optimal rank, the calculated and estimated experimental errors should be approximately equal.[10]

In general terms, a plot of the error between the regenerated and raw spectra *versus* the rank (number of included factors) should initially decrease as significant factors are included. Once the optimal rank has been reached, the plot should flatten out with only small subsequent variations according to the amount of error spanned by later noise factors. Factor-assessing methods using estimates of RE and XE are limited to cases where errors are relatively uniform and are not systematic or erratic. As mentioned above, they are also limited to systems where the experimental uncertainty can be well estimated, which can be difficult to achieve. Estimation of IE can be achieved empirically without prior estimation of the experimental errors, but is only useful if errors are uniformly and randomly distributed, in which case the IEs converge to a minimum value.[10] During this type of analysis it is worth remembering that IE and XE are always less than RE; it is the relationship IE $\leq$ RE that permits the noise removal properties of factor analysis.

There are some empirical functions such as the factor indicator function (IND) and autocorrelation (AUTO) functions,[10, 13] that can be used to determine the optimal rank. The advantage of the IND value, is that the IND *versus* rank plot has a minimum (rather than just flattening off) at the indicated optimal rank and has been demonstrated to work well when the error is random and uniformly distributed throughout the data matrix.[13] However, there are reservations about its use since the physical origins on which the function works are not well understood.[10] The AUTO function assesses the smoothness of the spectra and monitors the factors for rapid fluctuations of sign or intensity. This technique works poorly on spectra which have high signal to noise ratios or systematic errors.[10]

The most commonly used methods for determining the optimal rank are statistical. Some involve the analysis of variance (ANOVA) of the eigenvalues, where affects of including successive factors are compared using *F*-tests. A method of this type, originally developed by Malinowski, has been refined to use normalised or reduced eigenvalues (REV), which are less susceptible to variations due to changes in the degrees of freedom.[4, 10] Tests are generally performed comparing the last two eigenvalue first, and then working towards the first eigenvalue while pooling the results. When the correct number of factors is reached, the *F*-value calculated jumps over a significant value, usually 5-10 %, and all factors beyond this point are included in the basis set.[10, 14]

Another popular statistical method is selection by validation, which uses the predicted error sum of squares (PRESS) values, calculated according to equation (3.25), to estimate the error between the predicted ($\hat{c}_i$) and known ($\tilde{c}_i$) concentrations.

$$\text{PRESS} = \sum_{i=1}^{m} (\tilde{c}_i - \hat{c}_i)^2 \tag{3.25}$$

PRESS values can be calculated for a set of data that is separate from the data used in the calibration but contains similar spectra and is known as the test or validation set. In the absence of a validation set, samples can be removed one at a time from the calibration and their concentration can be predicted by using a calibration model generated from the remaining samples. When this is repeated for all samples in the calibration it is referred to as a "leave-one-out" cross validation or inner validation method.

PRESS values can be calculated as a function of the number of factors to determine the optimum rank in a model. Generally, the PRESS values initially decrease rather quickly as the number of factors increases, then flatten out before increasing slowly with the inclusion of more factors (Figure 3.4).[4, 10] Ideally, the model with the minimum PRESS value would represent the best fit to the data. But since PRESS values are calculated for a finite number of samples, they are subject to errors and, as a consequence, the rank that yields the lowest PRESS can lead to overfitting (too many factors).[9] So in practise, it is

normal to use the model that uses the fewest factors and has a PRESS that is not significantly different from the minimum value.[9]

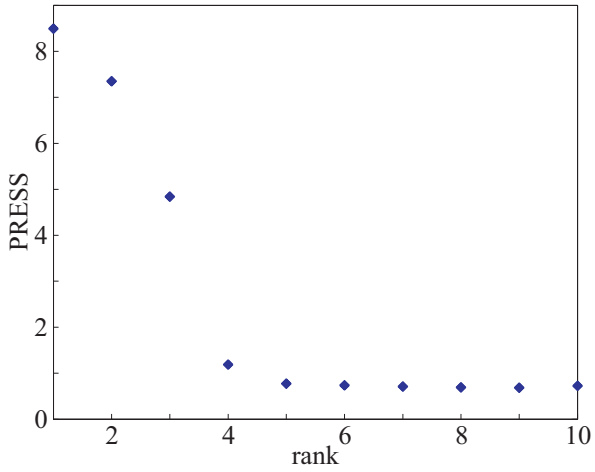


Figure 3.4: A typical PRESS *versus* rank plot. The rank with the smallest PRESS value is nine, but *F*-statistics tests and other empirical methods described in the text suggest a smaller rank of four or five.

The significance of the differences between PRESS values can be evaluated using an *F*-test method, outlined below (details can be found in the paper by Haaland and Thomas).[9] The model with the minimum PRESS value has the rank r^*. The PRESS values for all the models that contain fewer factors are ratioed with that for rank r^* to give an *F*-statistic as in equation (3.26).

$$F_r = \frac{\text{PRESS}_r}{\text{PRESS}_{r^*}} \qquad (3.26)$$

where $r = 1, 2, 3 \ldots r^*\text{-}1$

The optimum rank is determined to correspond to the model that has the smallest r (number of factors) but obeys the inequality $F_r < F_{\alpha;m,m}$, where $F_{\alpha;m,m}$ is the (1-α) percentile of the Snedecor's *F*-distribution with m (the degrees of freedom) being equal to the number of samples. The choice of significance level, α, is arbitrary but must be carefully considered. If it is too small, insufficient factors will be included in the calibration model and the data will be underfitted. If it is too large then too many factors will be included, leading to overfitting. Haaland and Thomas found that $\alpha = 0.25$ gave good results.[9]

If it is assumed that the prediction errors have a mean of zero, are mutually independent, and are normally distributed, then the probability expression in equation (3.27) holds, where σ_r^2 and σ_{r*}^2 are the predicted error variances for models with r and r* factors respectively and α is the selected probability value.

$$\text{Prob}\{F_r > F_{\alpha;m,m} | \sigma_r^2 = \sigma_{r*}^2\} = 2\alpha \quad (3.27)$$

In applications to real data, there are likely to be a positive correlation between the prediction errors of models with r and r* factors, and so equation (3.27) is modified to $\text{Prob}(F_r > F_{\alpha;m,m}) < 2\alpha$. A possible effect of this correlation is the underfitting of the data by not including enough factors.

An alternative method for determining the optimal rank is to select the model that has the lowest rank and whose PRESS is within one standard error of the minimum PRESS.[9]

Another method for determining the optimal rank is to choose the model with the minimum value of the Crit_r function, as defined for rank r in equations (3.28)–(3.31), which has been empirically shown to return reasonable optimal ranks.[15] This function adds to the PRESS value a term linearly dependent on the rank, which makes it easier to detect the minimum in the plot against r. It also attempts to overcome the problem of overfitting by the inclusion of too many factors, since there is a bias towards lower ranks when the PRESS values are similar. Two slightly different methods for determining s have been presented in equations (3.30) and (3.31). The simpler and less computationally intensive expression, shown in equation (3.30), is used by some computer programmes such as UNSCRAMBLER®.[15] Martens and Dardenne [15] used a modification of this approach shown in equation (3.31), where $\text{PRESS}_{r\,\text{max}}$ is limited by the number of samples in the calibration set.

$$\mathrm{Crit}_r = \mathrm{PRESS}_r + r\omega \tag{3.28}$$

$$\omega = 0.05s \tag{3.29}$$

$$s = \mathrm{PRESS}_0 \tag{3.30}$$

$$\text{or} \quad s = \sqrt{\mathrm{PRESS}_0^{\,2} + \mathrm{PRESS}_{r\,\max}^{\,2}} \tag{3.31}$$

Similar methods can be used for comparisons between models created by different methods; in this case the *F*-statistic is calculated using an equivalent to equation (3.32) that pertains to M1 and M2, corresponding to the different models.

$$F = \frac{\mathrm{PRESS_{M1}}}{\mathrm{PRESS_{M2}}} \tag{3.32}$$

Making the same assumptions used previously for the *F*-statistic, α is the probability that the null hypothesis (H_0) is rejected in favour of the alternative hypothesis (H_A) when the latter is in fact true. The null and alternative hypotheses are expressed in equations (3.33) and (3.34) respectively, and the alternative hypothesis is accepted when $F > F_{\alpha;m,m}$. Commonly chosen α values are 0.1 and 0.05.

$$H_0 : (\sigma_{M1}^{\,2} = \sigma_{M2}^{\,2}) \tag{3.33}$$

$$H_A : (\sigma_{M1}^{\,2} > \sigma_{M2}^{\,2}) \tag{3.34}$$

Since PRESS values depend on the number of samples, they are not useful for comparing calibrations with different numbers of samples. Instead, the root-mean-square errors (RMSE), as defined in equation (3.35), can be used. An RMSE determined for an independent validation set is called a root-mean-square error of prediction (RMSEP); that for cross validation is called a root-mean-square error of cross validation (RMSECV). RMSEs can be subjected to the same *F*-test procedures as PRESS values.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{m}(\tilde{c}_i - \hat{c}_i)^2}{m}} \tag{3.35}$$

3.9 Partial least squares

PLS was developed in the 1960s by Herman Wold for use in econometrics. It was first used for chemical applications in the 1970s by innovative scientists such as Svante Wold (Herman's son), Harald Martens and Bruce Kowalski. The following treatment of PLS is similar to that presented by Haaland and Thomas.[9]

PLS and PCR both involve factor analysis, but PLS goes further by circumventing the assumption (inherent in PCR) that the **C**-matrix is error free. It uses information in both the **C-** and **A**-matrices, simultaneously, to build a calibration.

There are two abstract spaces in PLS; the spectroscopic factor space and the concentration factor space. PLS effectively performs PCA on both matrices so that **A** and **C** are then represented as the sum of the inner product of a loading vector ($\mathbf{p}_i$ or $\mathbf{q}_i$) and a score vector ($\mathbf{t}_i$ and $\mathbf{u}_i$), as shown in equations (3.36) and (3.37), while $\mathbf{E}_\mathbf{A}$ and $\mathbf{E}_\mathbf{C}$ are the absorbance and concentration error matrices. It is then recognised that even though these score vectors have been calculated in different factor spaces, their changes after normalisation should be the same, since the fundamental information contained in both **A-** and **C**-matrices is the same. This statement is true as long as changes in absorbance are perfectly linear with changes in concentration. Therefore achieving the maximum correlation between the two score vectors is an important aim of PLS.

$$\mathbf{A} = \sum_i \mathbf{t}_i \mathbf{p}_i + \mathbf{E}_\mathbf{A}$$

$$= \mathbf{TP} + \mathbf{E}_\mathbf{A} \quad (3.36)$$

$$\mathbf{C} = \sum_i \mathbf{u}_i \mathbf{q}_i + \mathbf{E}_\mathbf{C}$$

$$= \mathbf{UQ} + \mathbf{E}_\mathbf{C} \quad (3.37)$$

The changes in the two score vectors are correlated since the information in **A-** and **C**-matrices are correlated. If the absorbance changes are perfectly linear with concentration changes then the scores should be identical. But this will not be the case due to noise in both **A** and **C.** The intention of PLS is to maximise the correlation between the $\mathbf{t}_i$ and $\mathbf{u}_i$ score vectors by a method that can be envisaged as a series of simplified CLS and ILS steps.

There are two types of PLS procedures, PLS-1 and PLS-2, (PLS-1 is the more widely used and is usually referred to as just PLS, although, definitions of these procedures can vary with different software packages). The difference between them lies with the treatment of the **C**-matrix. PLS-1 develops a calibration for each analyte in **C** individually, using one column of **C** at a time. PLS-2, sometimes referred to as global PLS, does not separate the columns of **C** but creates a model for all components in **C** simultaneously. The description of PLS-1 that is presented here is based on a modified nonlinear iterative partial least squares algorithm (NIPALS; the standard algorithm for computing partial least squares regression components or factors), where concentration information is directly introduced into the calibration.[9, 10] In practise, the NIPALS algorithm has been superseded by faster alternatives, such as SIMPLS and the "kernel" PLS,[16-18] but these give the same results.

The first step is evaluation of the **A**-matrix using an estimated loading vector, $\mathbf{w_1}$ (where the subscript 1 indicates the first PLS iteration), and vector **c**, taken as a column of **C** that

contains the concentration information for the component of interest (equation (3.38)), $\mathbf{w}_i$ is referred to as the "weight loading" because it is a weighting vector for **c**.

$$\mathbf{A} = \mathbf{c}\mathbf{w}_i + \mathbf{E}_\mathbf{A} \tag{3.38}$$

Equation (3.38) can be considered in a CLS sense by regarding $\mathbf{w}_i$ as the equivalent of calibration matrix **K** in equation (3.2). If the spectral changes were perfectly reflected in the concentration changes, then **c** could be considered a score vector of **A.** This is not usually the case, but if errors occur in **A**, noise can be removed by using **c**, which is correlated to **A** but free of spectroscopic noise. If this treatment is performed on a data set that has errors in a few spectra, the result is that the estimated calibration is closer to the ideal calibration and the results are not as badly affected by the poor samples.

The outline above shows how PLS can improve a calibration by removing noise and producing a calibration that is closer to the ideal error-free situation. To do this, the first step is to find an initial weight loading vector, $\mathbf{w}_i$; this is carried out by calculating the least-squares solution of equation (3.39) (equation (3.39)), where $\mathbf{w}_i$ is similar to a PCA loading vector.

$$\mathbf{w}_i = \frac{\mathbf{A}_i^{\mathbf{T}}\mathbf{c}}{\mathbf{c}^{\mathbf{T}}\mathbf{c}} \tag{3.39}$$

Where $\mathbf{A_1} = \mathbf{A}$

$$\hat{\mathbf{w}}_i = \frac{\mathbf{w}_i}{(\mathbf{w}_i^{\mathbf{T}}\mathbf{w}_i)^{0.5}} \tag{3.40}$$

The next step is to normalise $\mathbf{w}_i$, as shown in equation (3.40), to give a set of mutually orthogonal $\hat{\mathbf{w}}_i$ vectors. The normalised weight vector, $\hat{\mathbf{w}}_\mathbf{1}$, can be considered to be the first-order, least-squares estimate of the pure analyte spectrum for the **c** component. The next step is very similar to CLS prediction where a projection of **A** onto $\hat{\mathbf{w}}_i$ is then calculated to give an absorbance score $\mathbf{t}_i$, equation (3.41). The score vector $\mathbf{t}_i$, represents the intensity of the weight loading vector $\hat{\mathbf{w}}_i$ in the calibration samples for the new coordinate system. Another way of looking at this is that if $\hat{\mathbf{w}}_\mathbf{1}$ is the least squares

estimate of the pure analyte spectrum, then $\mathbf{t_1}$ is the first order estimate of the amount of the pure analyte present in each calibration spectrum. This leads to the major difference between PLS and PCR, since $\mathbf{t}_i$ is related to both **A** and **c**.

$$\begin{aligned} \mathbf{A} &= \mathbf{t}_i\hat{\mathbf{w}}_i \\ \mathbf{t}_i &= \frac{\mathbf{A}\hat{\mathbf{w}}_i^{\mathbf{T}}}{\hat{\mathbf{w}}_i^{\mathbf{T}}\hat{\mathbf{w}}_i} \\ &= \mathbf{A}_i\hat{\mathbf{w}}_i \end{aligned} \tag{3.41}$$

Once $\mathbf{t}_i$ has been calculated, its intensity in the PLS coordinated system needs to be related to the analyte concentration. This can be carried out using ILS methodology (equation (3.42)), where the concentration loading scalar, q_i, is determined according to equation (3.43). This process is similar to PCR except that the relationship between $\mathbf{t}_i$ and $\mathbf{c}_i$ (or concentration residuals) is found after each $\hat{\mathbf{w}}_i$ has been estimated.

$$\mathbf{c} = \mathbf{t}_i q_i \tag{3.42}$$

$$q_i = \frac{\mathbf{t}_i^{\mathbf{T}}\mathbf{c}}{\mathbf{t}_i^{\mathbf{T}}\mathbf{t}_i} \tag{3.43}$$

Since q_i is a scalar and $\mathbf{t}_i$ is proportional to **c** through equation (3.37), **c** must be proportional to $\mathbf{u}_i$ which implies that $\mathbf{t}_i = \mathbf{u}_i$. To avoid the collinearity problems that can present in ILS, it is important that the $\mathbf{t}_i$ vectors are orthogonal. Orthogonality can be obtained by using a new model for **A** based on the latent vector $\mathbf{t}_i$, equation (3.45), where $\mathbf{p}_i$ is the PLS loading vector. The PLS absorbance loading vector $\mathbf{p}_i$ is then calculated using least squares regression of equation (3.44).

$$\mathbf{p}_i = \frac{\mathbf{A}^{\mathbf{T}}\mathbf{t}_i}{\mathbf{t}_i^{\mathbf{T}}\mathbf{t}_i} \tag{3.44}$$

The first PLS loading vector, $\mathbf{p_1}$, is different from the first PCA loading vector, which spans the maximum variance in **A**. Instead, it attempts to span the maximum variance in **A** while simultaneously correlating with $\mathbf{t}_i$, which, in turn, approximates **c**. It should also

be noted that, unlike PCA loading factors, the $\mathbf{p}_i$ loading vectors are not mutually orthogonal.

The next step is to set up **A** and **c** for the calculation of the next factor. This is performed by computing the spectral and concentration residuals in **A** and **c** that remain unmodelled by $\mathbf{w}_i$ (equations (3.45) and (3.46)).

$$\mathbf{A}_{i+1} = \mathbf{A}_i - \mathbf{t}_i\mathbf{p}_i \tag{3.45}$$

where $\mathbf{A}_1 = \mathbf{A}$

$$\mathbf{c}_{i+1} = \mathbf{c}_i - \mathbf{t}_i q_i \tag{3.46}$$

where $\mathbf{c}_1 = \mathbf{c}$

$\mathbf{A}_{i+1}$ and $\mathbf{c}_{i+1}$ can be considered in terms of $\mathbf{E}_\mathbf{A}$ and $\mathbf{e}_\mathbf{c}$, which are the error matrices or vectors that can be fed back into the model in order to calculate more factors and to better fit the calibration data.

Once the PLS procedure has been repeated a sufficient number of times so that there are enough factors to adequately model **A** and **c**, the system is reduced to equations (3.47) and (3.48). These are similar to the original set of PLS equations (equations (3.36) and (3.37)), except that there is a score vector, $\mathbf{t}_i$, common to both and the model is only representative of one analyte concentration (i.e. one column of **C**). The optimum rank is determined in ways similar to that outlined for PCR in Section 3.8 . The methods most widely used in PLS are statistical functions using PRESS or RMSE values.

$$\mathbf{A} = \sum_i \mathbf{t}_i\mathbf{p}_i + \mathbf{E}_\mathbf{A} \tag{3.47}$$

$$\mathbf{c} = \sum_i \mathbf{t}_i q_i + \mathbf{e}_\mathbf{c} \tag{3.48}$$

The first step of applying the PLS method is to apply any pre-treatments performed on the calibration spectra to the spectra of the unknowns. These spectra are then collected in a matrix $\mathbf{A}_\mathbf{unk}$, to which the first weight loading vector, $\hat{\mathbf{w}}_i$, is applied to determined the

score vector, $\mathbf{t}_i$, (equation (3.49)). The score vector is then used to predict analyte concentration according to equation (3.50).

$$\mathbf{t}_i = \mathbf{A}_{\mathbf{unk}}\hat{\mathbf{w}}_i \tag{3.49}$$

$$\hat{\mathbf{c}}_i = \mathbf{t}_i q_i \tag{3.50}$$

The absorbance residuals calculated using equation (3.45) are then used in equation (3.49) and the process is repeated for all factors included in the calibration. The contributions to the concentration calculated in equation (3.50) for the analyte of interest, are summed to determine the total concentration in the unknown sample.

An alternative way of performing a PLS prediction is to calculate a regression vector, $\boldsymbol{\beta}$, equation (3.51), where $\mathbf{W}$ and $\mathbf{P}$ have r rows that contain the $\hat{\mathbf{w}}_i$ and $\mathbf{p}_i$ vectors, respectively. The regression vector can then be multiplied by the unknown sample spectrum to predict the analyte concentration as in equation (3.52). The disadvantage of this technique is that the residual spectrum, which could have been used to determine how well the model fitted the unknown spectrum (and hence how well the model had performed), cannot be evaluated.[9]

$$\boldsymbol{\beta} = \mathbf{W}(\mathbf{P}^{\mathrm{T}}\mathbf{W})^{-1}q_i \tag{3.51}$$

$$\hat{\mathbf{c}}_i = \mathbf{A}_{\mathbf{unk}}\boldsymbol{\beta} \tag{3.52}$$

In PLS-2, a calibration is built for two or more analytes simultaneously, with the concentration vector, $\mathbf{c}$, being replaced by a concentration matrix, $\mathbf{C}$. This leads to a few changes in relation to the PLS-1 algorithm. Vector $\mathbf{u}$ is substituted for $\mathbf{c}$ in equation (3.39), with the initial estimate for $\mathbf{u}$ being a column of $\mathbf{C}$; and two score vectors are determined, one each for the concentration factor space and spectroscopic factor space. The substitution of $\mathbf{C}$ into equation (3.42) also causes changes, with q_i becoming $\mathbf{q}_i$, which is a concentration loading vector rather than a scalar. These changes lead to an iterative cycle in the PLS-2 algorithm with the aim to obtain maximum correlation between the score vectors $\mathbf{u}_i$ and $\mathbf{t}_i$. This is achieved by the inner relationship, $\mathbf{b}_i$, defined in equation (3.53). When convergence occurs, $\mathbf{b}_i$ is close to a unit matrix.

$$\mathbf{u}_i = \mathbf{b}_i\mathbf{t}_i \quad (3.53)$$

Once convergence of the score vectors has been achieved, the PLS loading vector, $\mathbf{p}_i$, is calculated. The **A** and **C** residual matrices are then calculated in a similar fashion to PLS-1. The next PLS factor is then calculated using the same procedure as in PLS-1, except with the new **A-** and **C**-matrices. Once the process has been repeated for all factors, the system can be expressed as equations (3.36) and (3.37).

The results obtained from PLS-1 and PLS-2 are not the same. PLS-1 generally has better predictive properties, due to a number of reasons.[9, 10, 19] The most significant is that PLS-1 allows different analytes to have different optimal ranks whereas PLS-2 is restricted to a single optimal rank for all analytes.[9, 10] Also, in PLS-2, the individual analyte concentrations have to be corrected for differences in their precision, which, in most cases is unknown.[9] If analyte concentrations are correlated in both the calibration and unknown samples, then PLS-2 can have better predictive properties than PLS-1.[19] Building on this, if one analyte concentration is known inaccurately but is strongly correlated to another, accurately known analyte concentration, PLS-2 can be used to generate a more robust calibration.[10]

When many components are to be quantified, PLS-1 requires more data storage space than PLS-2.[19] But the interpretation of the information contained in PLS-2 score and loading vectors is more difficult, since the parts relating to the individual analytes becomes intimately mixed;[9] PLS-1 weight loading vectors can be useful in making spectral band assignments.

3.10 Comparison of partial least squares to the other modelling methods

PLS is a full spectrum method (like PCR and CLS) and a factor analysis method (like PCR), which utilises the inverse Beer's law relationship expressed in ILS. ILS, PLS and PCR can be used not only to estimate analyte concentrations but also to estimate chemical and physical properties from spectra.[9] This is possible when the physical

property information, rather than concentrations, is used in the calibration process, but can not be achieved using CLS since all spectral intensities need to be assigned in the calibration process. Since PCA and PLS can both remove spectral noise in the compression of **A**, the assumption that the dominant errors are in the concentrations rather than the spectra is more closely approximated than it is in ILS.[9]

PLS is similar to PCR in that both are soft modelling methods, meaning that they can account for some types of non-linearities, particularly those that can be described using additive linear models. For example molecular bonding changes that lead to adduct formation involving intermediates can generally be considered to cause non-linearities in a Beer's law sense since no explicit information is known on concentrations of the intermediates and/or adducts.[9] Such systems can be more accurately fitted by PLS and PCR, than hard modelling techniques like CLS, but some molecule interaction information is lost from the residuals. In this way, the qualitative information extracted from PLS or PCR is less than that from CLS. It has also been observed that PCR is less useful at obtaining qualitative information than PLS.[9]

Because the PCR basis set is calculated using only the spectral data, the regenerated spectra are likely to fit the calibration data better than a PLS model; but PLS is likely to make better predictions on unknown samples.[11, 19] When comparing the ways in which PLS and PCR calibrations handle errors, it has been discovered that PLS deals better with random errors, such as instrument noise, whereas PCR deals better with systematic noise, such as detector saturation; therefore it is important to evaluate the types of errors present in the system before applying a model.[10] Also, since PLS uses both concentration and spectral information, it is more disposed than PCR to evaluate a stable calibration for minute components in the presence of intense signals since information about these components is not likely to be present in the first eigenvectors used in the basis set.[10]

Thomas and Haaland compared the four calibration methods described above and found that PLS was the best, or almost the best, performer for a wide range of conditions.[11] The results from PLS and PCR calibrations were very similar, the major difference being that

PLS out-performed PCR when random linear baselines had been added to the calibration spectra and when the analytes of interest had strongly overlapping features. The superior performance of PLS was attributed to the inclusion of concentration information in the modelling process. PCR did have better predictive properties than PLS when there were relatively large errors in the concentration data. Again, this was attributed to the concentration data entering the PLS model at two points (spectral decomposition and regression) rather than once in PCR.

3.11 Outlier detection and error analysis for PLS

A very important part of the modelling process is the evaluation of model performance. If a calibration is to be used, the user needs to know the conditions and ranges over which it is applicable. There should also be some sort of uncertainty determined for the results. Error analysis methods for univariate methods are well established and are defined in terms of readily calculated detection limits and confidence internals.[20] Unfortunately, the equivalent processes are far more complicated for multivariate methods.

Once a calibration model has been created, its quality can be checked by applying it to samples in the calibration set or separate validation (or verification) data sets if they are available. The residuals for both the concentration and spectral components can be analysed to determine if any data are sufficiently unusual to be regarded as outliers that should be removed from the data set or subjected to further analysis. Outliers are generally defined as data that are not well modelled by the calibration. This may be because: they contain unexpected, unknown components that have not been included in the calibration; their concentration values have large errors; the spectra are very noisy in comparison to the calibration spectra; or the conditions lie outside the calibration range.[19]

Residuals are the differences between the original measured spectral and concentration information used in the calibration and those regenerated by the calibration. The matrices corresponding to the absorbance and concentration residuals, respectively, are the $\mathbf{E}_A$ and

$\mathbf{E_C}$ found in equations (3.47) and (3.48). *F*-statistics can be used to decide whether a sample has a residual that is statistically different from those of other samples in the calibration or validation sets. For concentration residuals, the *F*-statistics are calculated using equation (3.54), where *m*-1 are the degrees of freedom and e_{cj} is the concentration residual for the *j*th sample. Spectral *F*-statistics are calculated using equation (3.55), where $\mathbf{a}_j$ is the spectrum of the sample left out during cross validation, *x* and *y* are the degrees of freedom estimated by equations (3.56) and (3.57),[9, 21] and *n* is the number of spectral frequencies included in the analysis and found in $\mathbf{a}_j$.

$$F_{1,m\text{-}1}c_j = \frac{(m\text{-}1)(e_{cj}^2)}{\sum_{i\neq j} e_{ci}^2} \tag{3.54}$$

$$F_{x,y}\mathbf{a_j} = \frac{(m\text{-}1)\sum_{k=1}^{n} e_{ajk}^2}{\sum_{i\neq j}\sum_{k=1}^{n} e_{aik}^2} \tag{3.55}$$

Where

$$x = \frac{(n - r)}{2} \tag{3.56}$$

$$y = \frac{(n - r)(m - r - 1)}{2} \tag{3.57}$$

Using equation (3.55), the *F*-statistics can be calculated for the spectral residuals for unknown samples whose concentrations are to be predicted by the calibration model; $\mathbf{a}_j$ is replaced by $\mathbf{a_{unk}}$, the unknown sample spectrum, and e_{ajk} is replaced by e_{aunkk}, the unknown sample residual or error, and the degrees of freedom are changed from *m*-1 to *m*. *F*-statistics calculated from unknown samples can be used as a way to detect troublesome samples whose predicted concentrations are likely to have a high degree of associated error.

An interesting approach that has been applied to multivariate analysis is to combine target factor analysis (TFA) with *F*-statistics.[14] A vector of interest, such as a pure spectrum of a suspected analyte, is tested to see whether it lies within the factor space defined by the model, with *F*-statistics being used to determine the level of significance

of any discrepancy. A general rule is that if the calculated α value is less than 5-10% then the pure spectrum is considered not to lie significantly outside the factor space and therefore the corresponding analyte *is* present in the calibration samples. If the α value is greater, then the spectrum is not considered to be part of the factor space and therefore the samples do not contain the suspected analyte.

As mentioned earlier in this section, it is important that results from a model are accompanied by a measure of their uncertainty,[20] despite the difficulties of applying univariate concepts such as standard errors of prediction, prediction intervals, and limits of detection or discrimination. Since PLS model has become the default chemometric model there has been a lot of research in this field in the last decade.[20, 22-24]

Mean squared error (MSE) and RMSE values calculated in the process of determining the optimal rank and evaluating a model's performance can be used to estimate model error; for example MSEP and RMSEP can be used as estimates of model variance and standard deviation. But problems arise when they are used to estimate prediction uncertainty since they summarise information intrinsic to the calibration set rather than to the unknown samples to which the calibration is to be applied. First amongst these problems is the fact that they are single values taken to be representative of all concentrations. But the uncertainty associated with any given concentration will be dependent on many factors, some quite specific to the sample. Further, as a general principle, samples near the centre of the calibration are likely to have smaller uncertainties than those on the outer edges of the calibration range, since the confidence bands are not linear (see Figure 1 in Olivier *et al*.[20]). Secondly, the apparent RMSE values are affected by errors in the reference method by which the concentrations of the calibration and validation samples are established. What are really desired are the values based on error free concentrations.[22, 24]

Although estimates of standard errors obtained using apparent RMSE values are biased, they are commonly used in the PLS and PCR literature due to the difficulties of determining more scrupulous measures of model uncertainty. When they are used, it is generally thought to be better to use a validation set that is independent of the calibration

set than to rely on a cross-validation.[4] This may be true in the case of large data sets, but is not necessarily true for small data sets, as demonstrated by Martens and Dardenne.[15] These researchers analysed a large set of near-infrared spectra collected from maize extracts to determine protein content. Smaller data sets selected from the pool of spectra were processed using different methods to determine RMSE values, with sufficient repetition to make comparisons statistically valid. Four RMSE values were determined: from cross validation, cross verification, set test validation and test set verification.[‡‡] They concluded that the use of an independent validation set was wasteful and that it generally gave an over-optimistic estimation of future predictive error. Overall it was shown to underestimate the resulting "true apparent" prediction error. An independent verification set gave better estimates of the true apparent prediction error, but individual estimates had a high degree of variability. Cross verification and cross validation gave the best results, although the estimated true apparent prediction error was slightly underestimated. It was concluded that for small data sets it was best to use cross validation since this gave the best estimate of the true apparent prediction error. Cross verification also gave good results although it required more calculation for little or no gain over cross validation. This result is supported by the results of Mevik and Cederkvist who compared the RMSEP values from PLS and PCR models using twelve different methods.[25] Their results indicated that the best estimates of RMSE were determined by leave-one-out cross validation and the "0.632 estimate", a heuristic modification of the bootstrap method (described later in this section).

A number of strategies have been developed to determine the actual RMSE from the apparent RMSE. Many of the papers offer statistically rigorous ways of determining sample-specific standard errors of prediction. But, unfortunately, they require knowledge of the noise for both the reference and unknown spectra, which are usually unknown and therefore make implementation difficult. This has led to the development of more "user friendly" techniques. One popular example is the error in variable (EIV) method summarised by equations (3.58) to (3.60). Here the $\mathbf{c}$ ($m \times 1$) vector contains the true and

[‡‡] Cross verification and test set verification are not explained here but an explanation of the methods can be found in reference 15. Martens, H. A. and Dardenne, P. Validation and verification of regression in small data sets, *Chemometrics and Intelligent Laboratory Systems*, **1998**, *44*, 99-121.

unobservable analyte concentration, **A** ($m \times n$) is the matrix of the true unobservable spectra, the measured and therefore observable values are denoted as $\tilde{\mathbf{c}}$ and $\tilde{\mathbf{A}}$, respectively, $\Delta\mathbf{c}$ and $\Delta\mathbf{A}$ contain the unobservable measurement errors in both **c** and **A**, respectively, and **β** is the regression vector from the PLS model calculated in equation (3.51).[24]

$$\mathbf{c} = \mathbf{A}\boldsymbol{\beta} + \mathbf{e} \tag{3.58}$$

$$\tilde{\mathbf{c}} = \mathbf{c} + \Delta\mathbf{c} \tag{3.59}$$

$$\tilde{\mathbf{A}} = \mathbf{A} + \Delta\mathbf{A} \tag{3.60}$$

The EIV expression has been used to derive a number of expressions for systems where there are different sources of uncertainty. Faber and Kowalski have published an expression that accounts for all measurement noise in the data,[26] but the application of it to real data has been hindered by the need for accurate reference values for all components, a condition that is rarely met.[22, 24] A simplified EIV model has been proposed that assumes that most of the error is in the reference concentration values rather than the spectra, and therefore ignores terms that become very small when the PLS calibration explains a substantial part of the variance in the analyte concentration.[24] This is expressed in another way in a second paper where the same expression is obtained by ignoring uncertainties in the PLS scores,[27] although both papers also assume that noise present in the calibration and prediction data is similar and independently distributed. The resulting expression for standard error of prediction, σ_{PE}, is shown in equation (3.61), where $\|\boldsymbol{\beta}\|$ is the Euclidean norm of the PLS regression vector, $\sigma_{\Delta c}^2$, $\sigma_{\Delta A}^2$ and σ_e^2 represent the variances of the measurement errors in the reference methods, spectra, and the concentration residual. *h* is the unknown sample's leverage, calculated using equation (3.62), where $\mathbf{t}_{\mathbf{unk}}$ is the PLS score vector for the unknown sample and **T** is the ($m \times r$) matrix from the PLS calibration model. Equation (3.61) can be modified for mean centred data by the adding $1/m$ to *h*.

$$\sigma_{PE} = \sqrt{h(\|\boldsymbol{\beta}\|^2 \sigma_{\Delta A}^2 + \sigma_e^2 + \sigma_{\Delta c}^2) + \|\boldsymbol{\beta}\|^2 \sigma_{\Delta A}^2 + \sigma_e^2} \quad (3.61)$$

$$h = \mathbf{t}_{\mathbf{unk}}^T (\mathbf{T}^T\mathbf{T})^{-1} \mathbf{t}_{\mathbf{unk}} \quad (3.62)$$

The first $(h(\|\boldsymbol{\beta}\|^2 \sigma_{\Delta A}^2 + \sigma_e^2 + \sigma_{\Delta c}^2))$ term in equation (3.61) is the model-dependent contribution from the calibration step. It can be interpreted in terms of the proximity of an unknown sample to the average calibration sample in the *r*-dimensional space of the PLS calibration model; the closer the unknown sample is to the calibration average, the smaller its leverage. Since the estimated concentration of a sample with a small leverage should be more precise than one further from the average calibration sample, the smaller the leverage, the smaller the contribution from the model error. The last two terms $(\|\boldsymbol{\beta}\|^2 \sigma_{\Delta A}^2 + \sigma_e^2)$ in equation (3.61), are the contributions from the unknown sample in the prediction step. Unlike the first term, these should be regression-method independent; however values for $\boldsymbol{\beta}$ need to be estimated using methods that are consistent with the model and therefore, in practise, they are also model, and hence method, dependent.

In a PLS model the selection of the optimal rank can be seen as a trade off between variance and bias. If too few factors are included, a prediction bias is introduced; if too many factors are included, then more variance is introduced.[27] If the model estimator of the mean squared error of calibration (MSEC) is assumed to account only for variance and is free of bias, then equation (3.61) can be simplified to equation (3.63). Again if the data are mean centred, then $1/m$ can be added to the leverage.

$$\sigma_{PE} = \sqrt{(1 + h)\text{MSEC} - \sigma_{\Delta c}^2} \quad (3.63)$$

$$\text{MSEC} = \frac{\sum_{i=1}^{m} (\tilde{c}_i - \hat{c}_i)^2}{m - \rho} \quad (3.64)$$

MSEC can be calculated using equation (3.64), where $\tilde{c}_i$ and $\hat{c}_i$ terms refer to the observed and predicted concentration values for the ith sample, respectively, m is the number of samples and ρ is the degrees of freedom used by the model. For CLS and PCR, ρ is equal to the model rank, but this is inappropriate for PLS where both spectral and concentration information are used.[24] It has been shown that use of the model RMSECV value in place of the RMSE value in equation (3.63) is a good alternative in PLS.[24] By substituting MSEC or RMSECV expressions into equation (3.61) the only variance that needs to be estimated is $\sigma_{\Delta c}^2$, which relates to the concentration error of the reference method in equation (3.63). In this expression, the variance is homoscedastic; that is, independent of sample concentration. This is not usually the case. More commonly, errors are heteroscedastic and a function of sample concentration. This can be factored into equation (3.63) by substituting $f(\hat{c}_i)$ for $\sigma_{\Delta c}^2$. The estimation of $f(\hat{c}_i)$ or $\sigma_{\Delta c}^2$ needs to be performed carefully because, if it is too pessimistic, the resulting estimate of the standard error of prediction will be overly optimistic. If there is no good estimate of $f(\hat{c}_i)$ or $\sigma_{\Delta c}^2$ it is safer to assume a value of zero and have a pessimistic estimator of standard error of prediction rather than an optimistic estimator that could lead to incorrect conclusions. An interesting result from this technique is that it is possible to have sample-specific uncertainty estimates that are smaller than the apparent RMSEP of the calibration model; that is to say, the calibration model can be used to predict the concentration of an unknown sample more precisely than the reference values used in the calibration. It should be noted that this is only possible if the variance in the reference concentrations is known.[24, 28]

Equation (3.63) describes a sample-specific standard error of prediction that should be used only for samples that are similar to the calibration samples in the sense that they do not contain interferences that are absent from the calibration samples. If it is not known whether a sample is similar, it is assumed that large spectral residuals (those that are not well fitted by the model) indicate unmodelled interferences and hence that equations (3.61) to (3.64) should not be used.

Other methods have been suggested that do not involve noise estimations, but which attempt to determine the standard error of prediction from information provided solely from the calibration model. One popular method is used in the UNSCRAMBLER® software package.[24] A number of papers have compared the results obtained by UNSCRAMBLER® and other more reliable methods with the conclusion that, although it is easy to implement, the results obtained are relatively poor.[24]

There are methods that are not theoretically based and can be used to estimate the uncertainty in prediction. Two, similar, so-called "re-sampling" methods are bootstrapping and noise addition. These are favoured in chemometrics since they entail few assumptions and can be used on systems where the deduction of statistical estimates is difficult. The basic concept behind them is random sampling with replacement, repeated many times to estimate statistical parameters. They generally provide good estimates of model parameters if the sample data are representative of the underlying population. Their major drawbacks are that they are computationally intensive and offer no insight regarding the sources of the prediction errors. In many previously mentioned publications, re-sampling methods have been used to determine the "true" result, which is then used in the comparison of other estimation techniques.[22, 24, 28]

A general bootstrapping algorithm has been described by Fernandez Pierna *et al.* and is summarised as follows.[24] A PLS calibration is built and the concentration residuals (e_1, e_2,…, e_m) for the samples are computed. The bootstrap vector, $\mathbf{e}_b^*$, is created from the pool of concentration residuals by random selection of m residuals with replacement (that is any one residual can be chosen any number of times). This bootstrap sample is then added to the predicted concentration vector, $\hat{\mathbf{c}}$, in equation (3.65).

$$\mathbf{c}_b^* = \hat{\mathbf{c}} + \mathbf{e}_b^* \tag{3.65}$$

The PLS model is then rebuilt using $\widetilde{\mathbf{A}}$ and $\boldsymbol{c}_b^*$ in place of $\widetilde{\mathbf{A}}$ and $\widetilde{\mathbf{c}}$. The predicted regression coefficient for the b^{th} bootstrap sample, $\hat{\boldsymbol{\beta}}_b$, is determined for the model with the same number of factors as in the original PLS calibration. Using this predicted

regression vector, the concentration for the unknown sample, $\tilde{\mathbf{c}}_{\mathbf{unk}}$, can be predicted as in equation (3.66).

$$\hat{c}_b = \tilde{\mathbf{c}}_{\mathbf{unk}}^{\mathrm{T}} \hat{\boldsymbol{\beta}}_b \tag{3.66}$$

These predicted concentration values are stored and the process is repeated B times, where B is the number of bootstrap samples. The standard deviation of the predicted concentration obtained can be used as an estimate of the sample-specific standard error of prediction. For these techniques to work well, the number of bootstrap iterations must be large enough that results are statistically significant. Typically this requires $B \approx 1000$ for one hundred or fewer samples.

Noise addition is achieved using a similar method to bootstrapping except that instead of creating bootstrap samples from the residuals, normally distributed noise is added to create a noise sample. The remainder of the process is similar to the bootstrapping,[24] and the results are similar, although is has been suggested that noise addition is more versatile since it can deal with correlated noise.[22]

There are other techniques that have been used to estimate prediction uncertainties of multivariate calibrations. One of these was suggested by Lorber and involves the estimation of the "net analyte signal".[29] It assumes a linear additive model, like the one used for CLS in equation (3.1). Lorber defines the net analyte signal of a component to be the part of an analyte's spectrum that is orthogonal to the spectra of all other components present in the calibration standards.[23] This is a useful approach since the error is related to the net contribution of each component to the spectrum rather than the global response. Once the net analytical response of a component has been determined, detection limits can be established by using the same methodology as for univariate systems. Since this method uses a classical Beer's-law expression, there are problems with its application since it requires all components in the system to be known. This is difficult in many practical situations, in the same way that it is for the application of CLS. But even though it can be difficult to implement, its advent is considered to be an important advance in multivariate analysis because it was the first method to address the issue of multivariate

detection limits, and because it introduced the concept of net analytical signal. Further development of this technique has made the concept of net analytical signal applicable to inverse calibration models like ILS, where all the component and interference information is not needed.[23] Developments to the net analyte signal approach mean it is more easily implemented when not all of the information of the spectral component is known, although when applying the technique in this form the method requires prior assessment of the contributions of background noise and interfering substances. The extension of the original technique to inverse Beer's law type relationships does improve the ease with which this technique can be applied. However, it should be noted that the unknown samples to which the model is being applied should be similar to the calibration samples so they should not contain any interferences which are not modelled by the calibration samples.

Another technique, developed by Lorber and Kowalski, and which preceded the more user friendly techniques described earlier in this section, involves the estimation of confidence intervals for the inverse calibration model.[23] An advantage of this method is that it uses the errors of concentrations to estimate the limit of detection, therefore circumventing the difficulties of determining the net analytical signal (which would require knowledge about the errors in the regression matrix). Unfortunately this advantage is accompanied by a requirement for knowledge about the errors in spectral responses, which requires several assumptions that lead to a number of limitations. The first of these is that it is assumed that there is a relationship between the variance of the prediction concentrations and the variance of the measured concentrations of the calibration samples. There is no reason to assume this relationship, but it does allow the variance of the prediction concentrations to be estimated. Another limitation relates to the information content of the spectra. If there are strongly overlapping features or the experiment is badly designed, then colinearity issues are likely, which cause high prediction errors. The confidence interval is generally obtained by summing partial intervals from the different sources of error. The variances from the partial intervals are estimates of error in the unknown spectrum and error from the calibration standard concentrations. The limit of detection for an analyte in a specific sample can be

determined by deducing the confidence interval and checking whether it includes zero, if the confidence interval for a particular sample contains zero then the concentration for that sample is below the detection limit. It should be noted that the degree of confidence expressed for confidence intervals calculated used the above assumptions and techniques only considered type I (false positive) errors with no consideration of type II (false negative) errors. The general model can be modified to include different expressions for the variance in the unknown spectrum and also to ignore errors in the spectral responses.[23] Even with these changes there are still issues with independence of variables, normal distributions and homoscedasticity.

The final technique noted here is of limited application and involves the transformation of a multivariate model to univariate form, followed by application of univariate analysis.[20, 23] This approach is limited because of the constraints it puts on the data. For it to work there must be a very strong correlation between the data and the first factor, which therefore can be used to explain nearly all of the variance. For this to be true the collected spectra need to be specific to the analyte of interest and free of overlapping spectral peaks from interfering substances.

3.12 The methodology adopted in this thesis

In the systems analysed in this work, it is difficult to identify all components contributing to the spectra; interfering components are likely to include atmospheric gases (especially water vapour and carbon dioxide) and absorption by the fibre optic. These features are complicated by baseline variations and non-linearities that conspire to make explicit modelling methods (like CLS and ILS) difficult to implement. The multivariate models of PLS and PCR are more suited. As mentioned in Section 3.11 the PLS calibration method has become the default in chemometrics, and is the one adopted in this thesis. It offers the advantages over PCR of requiring fewer factors in the optimal model. Also, since there is a reasonable amount of error associated with the reference concentration data, the ability of PLS model to correlate information from the spectral and

concentration data blocks can be advantageous. Another appealing characteristic of PLS is that meaningful residuals can be determined and used to analyse the model performance if an appropriate algorithm is used in the calibration model; for the work presented in this thesis the spectral residuals for unknown samples could be determined.

The methods used to determine the optimal rank for a given model and to evaluate the performance of different models are statistically based. Due to the reasonable small number of samples in each calibration set, RMSECV (rather than RMSEP) values were determined and compared. The optimal ranks were investigated using *F*-statistics, Crit values, and visual inspection criteria, which generally gave consistent results and suggests a degree of robustness. RMSECV values were also used to compare different models built using the same data set.

A conservative approach to error analysis has been adopted, with RMSECV values used as estimates of standard errors of prediction. It is known that this is not the most accurate estimator since it ignores errors introduced by the reference data and heteroscedasticity. The data in this thesis have reference concentration errors that show various degrees of heteroscedasticity, but this is generally ignored in the analyses since the data sets span similar concentration ranges and have been prepared and characterised using the same reference method. Starting with equation (3.63) which assumes model variance can be modelled by MSEP, the σ_{PE} expression can be expanded to equation (3.67). If the leverage (h) is assumed to be small and the concentration error of the reference method, $\sigma_{\Delta c}$, is assumed to be negligible, then the prediction error expression can be reduced to a single term, $\sigma_{PE} \approx \sqrt{\text{MSEP}}$.

$$\sigma_{PE} = \sqrt{\text{MSEP} + h\text{MSEP} - \sigma_{\Delta c}^{2}} \qquad (3.67)$$

Under these conditions, and following IUPAC recommendations,[30] a detection limit can be determined using equation (3.68), where α is the probability of obtaining a false positive and ρ is the number of degrees of freedom associated with the RMSEP. Since, in the PLS algorithm, the score and loading vectors for each sample are not independent of

each other, it is very difficult to determine the effective degrees of freedom.[24] The situation is complicated further by the presence of several spectra per sample. To avoid being overly optimistic about the performance of the PLS calibration, a very conservative approach is adopted, throughout this thesis, that ρ can best approximated by $m - r$.

$$\begin{aligned} \text{DL} &= t_{1-\alpha,\ \rho} \times \sqrt{\text{MSEP}} \\ &= t_{1-\alpha,\ \rho} \times \text{RMSEP} \end{aligned} \tag{3.68}$$

If allowance is also made for the probability, β, of obtaining a false negative, then equation (3.68) is modified to equation (3.69).

$$\text{DL} = (t_{1-\alpha,\ \rho} + t_{1-\beta,\ \rho}) \times \text{RMSEP} \tag{3.69}$$

For samples that contain heteroscedastic errors in the reference loadings, this is a pessimistic estimation of the detection limit, which can be improved by using a RMSEP taken from a model that covers a narrower loading range.

Due to the nature in which many of the samples were prepared in the work presented in this thesis, surface loadings were commonly heterogeneous and, in most situations, multiple spectra were collected per sample. This led to three different data treatments, labelled NA:NA, NA:A, and A:A, being used for each set of data. In these designations, A stands for “averaged” and NA for “non-averaged”. For reasons that will become apparent, the NA:NA treatment is referred to as a leave-one-*spectrum*-out method, while the NA:A and A:A treatments are variations of leave-one-*sample*-out methods.

In the NA:NA treatment, the spectra obtained from all samples are considered independently throughout the calibration and cross validation process. In the NA:A treatment, the calibration is built by considering spectra from each sample independently. But the RMSE values are determined by cross validation in which all spectra from each sample in turn are removed together (hence leave-one-sample-out), their loadings are predicted from the remaining calibration spectra and the value is then averaged, (the same results are achieved if the PLS prediction is performed on the averaged *spectrum* from all

the spectra for that sample). Finally in the A:A treatment the averaged spectra are used throughout.

If the samples were ideal (perfectly homogeneous with exactly known loadings) there would be no difference between these three treatments. But heterogeneity and measurement uncertainties will cause them to give different results. Generally, the leave-one-spectrum-out, NA:NA, treatment is not recommended since spectra from the same sample are not independent. If one of them is removed in the cross validation process while the rest are retained in the calibration set, the overall procedure will be susceptible to bias. There are advantages and disadvantages for both of the two leave-one-sample-out techniques. In the NA:A treatment all of the information in all spectra is available to the PLS regression factors whereas averaging of the spectra in the A:A treatment means that some of this information is lost. On the other hand, if the surface loading is heterogeneous and only the average is known, the distribution of loadings returned from the spectra will overestimate the uncertainty of the PLS method, a problem that will be partly rectified by averaging the spectra to give a result closer to the actual average loading. If the samples have been "carefully" prepared the differences should be very small.

Leverage is a measure of the influence that a particular spectrum has over the PLS factors included in the calibration. A large leverage implies that a spectrum exerts a large influence over one or more of the PLS factors used in the calibration. A spectrum's leverage, h, in a PLS calibration can be calculated using equation (3.70), where $\mathbf{T}$ is the score matrix from the PLS calibration model and $\mathbf{t}_{cv}$ is the score vector for the spectrum calculated in an iteration of the cross validation process.

$$h = \mathbf{t}_{cv}^{T}(\mathbf{T}^{T}\mathbf{T})^{-1}\mathbf{t}_{cv} \tag{3.70}$$

A concentration residual for a spectrum is the difference between the predicted concentration (determined by the PLS calibration) and the "true" concentration (determined by the primary calibration method). These concentration residuals were

standardised as a fraction of the standard deviation for all spectra from all samples; those greater than ± 3 standard deviations were considered to be "large".

Possible statistical outliers were investigated by comparing the leverage for each spectrum with its concentration residual. Spectra that displayed both high leverages and high residual values were considered to be outliers and were removed from the calibration set along with the rest of the spectra from that sample. Spectra that displayed *either* high leverages *or* high residual values were highlighted for further analysis.

3.13 References

1. Wold, S. Chemometrics; what do we mean with it, and what do we want from it?, *Chemometrics and Intelligent Laboratory Systems*, **1995**, *30*, 109-115.
2. Danzer, K., Otto, M. and Currie, L. A. Guidelines for calibration in analytical chemistry part 2. Multispecies calibration (IUPAC Technical Report), *Pure and Applied Chemistry*, **2004**, *76*, 1215-1225.
3. Franke, J. E. In *Handbook of Vibrational Spectroscopy: Sample Characterization and Spectral Data Processing*; Chalmers, J. M. and Griffith, P. R.; John Wiley and Sons, LTD: Chichester, UK, 2002; Vol. 3, 2276-2292.
4. Kramer, R. *Chemometric Techniques for Quantitative Analysis*; Marcel Dekker Inc: New York, 1998.
5. Geladi, P. Some recent trends in the calibration literature, *Chemometrics and Intelligent Laboratory Systems*, **2002**, *60*, 211-224.
6. Mark, H. In *Handbook of Vibrational Spectroscopy: Sample Characterization and Spectral Data Processing*; Chalmers, J. M. and Griffith, P. R.; John Wiley and Sons, LTD: Chichester, UK, 2002; Vol. 3, 2258-2275.
7. Draper, N. R. and Smith, H. *Applied Regression Analysis*, 2nd ed.; John Wiley and Sons, Inc: New York, 1981.
8. Strang, G. *Linear Algebra and its Applications*, 3rd ed.; Harcourt Brace Jovanovich College Publishers: Orlando, Florida, 1988.

9. Haaland, D. M. and Thomas, E. V. Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information, *Analytical Chemistry*, **1988**, *60*, 1193-1202.

10. Hasegawa, T. In *Handbook of Vibrational Spectroscopy: Sample Characterization and Spectral Data Processing*; Chalmers, J. M. and Griffith, P. R.; John Wiley and Sons, LTD: Chichester, UK, 2002; Vol. 3, 2293-2311.

11. Thomas, E. V. and Haaland, D. M. Comparison of multivariate calibration methods for quantitative spectral analysis, *Analytical Chemistry*, **1990**, *62*, 1091-1099.

12. Malinowski, E. R. Theory of error in factor analysis, *Analytical Chemistry*, **1977**, *49*, 606-612.

13. Malinowski, E. R. Determination of number of factors and experimental error in a data matrix, *Analytical Chemistry*, **1977**, *49*, 612-617.

14. Malinowski, E. R. Statistical F-tests for abstract factor analysis and target testing, *Journal of Chemometrics*, **1988**, *3*, 49-60.

15. Martens, H. A. and Dardenne, P. Validation and verification of regression in small data sets, *Chemometrics and Intelligent Laboratory Systems*, **1998**, *44*, 99-121.

16. Dayal, B. S. and MacGregor, J. F. Improved PLS algorithms, *Journal of Chemometrics*, **1997**, *11*, 73-85.

17. de Jong, S. SIMPLS - an alternative approach to partial least squares regression, *Chemometrics and Intelligent Laboratory Systems*, **1993**, *18*, 251-263.

18. Lindgren, F., Geladi, P. and Wold, S. The kernel algorithm for PLS, *Journal of Chemometrics*, **1993**, *7*, 45-59.

19. Haaland, D. M. In *Practical Fourier Transform Infrared Spectroscopy: Industrial and Laboratory Chemical Analysis*; Ferraro, J. R. and Krishnan, K.; Academic Press, Inc.: San Diego, California, 1990.

20. Olivieri, A. C., Faber, N. K. M., Ferre, J., Boque, R., Kalivas, J. H. and Mark, H. Uncertainty estimation and figures of merit for multivariate calibration, *Pure and Applied Chemistry*, **2006**, *78*, 633-661.

21. Lindberg, W., Persson, J.-A. and Wold, S. Partial least-squares method for spectrofluorimetric analysis of mixtures of humic acid and ligninsulfonate, *Analytical Chemistry*, **1983**, *55*, 643-648.

22. Faber, N. M., Song, X. H. and Hopke, P. K. Sample-specific standard error of prediction for partial least squares regression, *Trends in Analytical Chemistry*, **2003**, *22*, 330-334.

23. Boque, R. and Rius, F. X. Multivariate detection limits estimators, *Chemometrics and Intelligent Laboratory Systems*, **1996**, *32*, 11-23.

24. Fernandez Pierna, J. A., Jin, L., Wahl, F., Faber, N. M. and Massart, D. L. Estimation of partial least squares regression prediction uncertainty when the reference values carry a sizeable measurement error, *Chemometrics and Intelligent Laboratory Systems*, **2003**, *65*, 281-291.

25. Mevik, B. H. and Cederkvist, H. R. Mean squared error of prediction (MSEP) estimates for principal component regression (PCR) and partial least squares regression (PLSR), *Journal of Chemometrics*, **2004**, *18*, 422-429.

26. Faber, K. and Kowalski, B. R. Propagation of measurement errors for the validation of predictions obtained by principal component regression and partial least squares, *Journal of Chemometrics*, **1997**, *11*, 181-238.

27. Faber, N. M. and Bro, R. Standard error of prediction for multiway PLS 1. Background and a simulation study, *Chemometrics and Intelligent Laboratory Systems*, **2002**, *61*, 133-149.

28. Faber, N. M., Schreutelkamp, F. H. and Vedder, H. W. Estimation of prediction uncertainty for a multivariate calibration model, *Spectroscopy Europe*, **2004**, *16*, 17-20.

29. Lorber, A. Error propagation and figures of merit for quantification by solving matrix equations, *Analytical Chemistry*, **1986**, *58*, 1167-1172.

30. Currie, L. A. Nomenclature in evaluation of analytical methods including detection and quantification capabilities., *Pure and Applied Chemistry*, **1995**, *67*, 1699-1723.

Chapter 4

Experimental

4.1 Instrumentation

IRRA spectra were collected using a prototype Remspec SpotView® grazing-angle head connected to the external beam port of a Bruker Vector 22 FTIR spectrometer by a ~1.5 m, 3 mm diameter, nineteen fibre chalcogenide-glass optical bundle. With the exception of a strong H-Se absorption band at ~2200 cm^{-1} (Figure 4.1), the fibre-optic transmits over the range from 5000 to 900 cm^{-1}. The grazing-angle head utilises off-axis parabolic mirrors to collimate the beam onto the sample (at a fixed incidence angle of 80° to the surface normal) and focus the reflected light to an integrated MCT detector, which is cooled with liquid nitrogen (Figure 4.1).[1] The FTIR spectrometer is controlled by a Micron Pentium 2 computer, which uses Bruker OPUS software to collect and analyse the spectra.

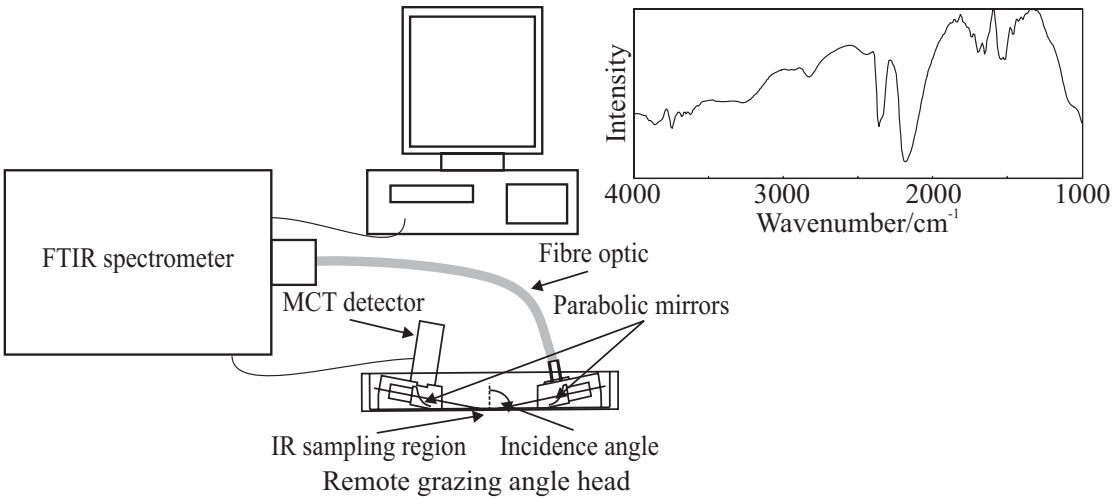


Figure 4.1: Schematic diagram of the instrument used for the measurement of the IRRA spectra from surfaces contaminated with organic compounds. The spectrum on the right shows the transmittance of the chalcogenide glass used in the optical-fibre bundle.

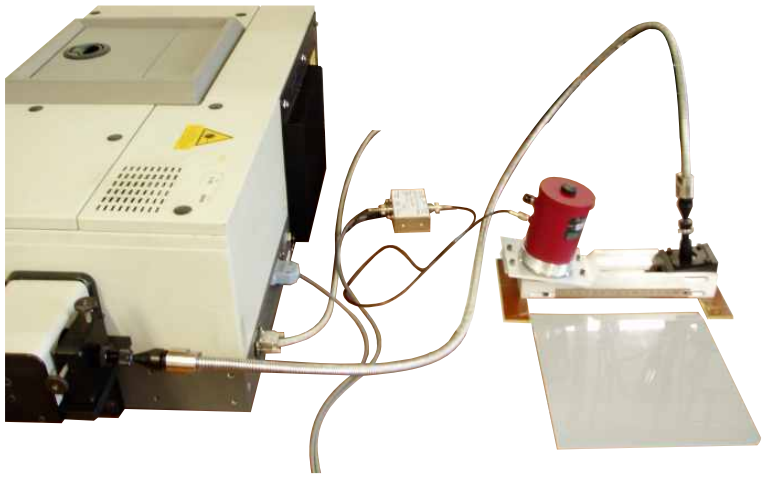


Figure 4.2: The photograph of the instrument showing the FTIR spectrometer and the sampling head with MCT detector; the fibre-optic cable is also visible.

The sampling head was then supported on blocks, visible in the photograph bottom left corner of Figure 4.2, which held it above the level of the bench. The substrate samples were placed on a clean Perspex tray and moved under the sampling head. The size of the tray is such that it could be rotated by 90° and still fit between the two supporting blocks. This allowed the IRRA spectra to be collected from a prepared sample surface with no adjustment of the grazing-angle head.

Spectra were collected over the range 1000–4000 cm^{-1} at 4 cm^{-1} resolution. A single-channel "background" spectrum of the clean surface was collected prior to sample deposition. The sample was then deposited (as discussed in Sections 4.6 to 4.8) and a single-channel "sample" spectrum was collected. The IRRA spectra were calculated according to equation (4.1);

$$RA = -\log_{10}\left(\frac{R}{R_0}\right) \tag{4.1}$$

where: RA = IRRA spectrum.

R_0 = background reflection-absorption intensity.

R = sample reflection-absorption intensity.

The number of spectra collected from each of the prepared samples was determined by plate size, as described in following chapters. Absorption spectra of the analytes

were collected using the internal sample compartment of the same FTIR spectrometer with its internal DTGS detector.

4.2 Infrared fibre optics

The most significantly innovative feature of the instrument is the use of mid-IR fibre optics to transmit the IR radiation from the spectrometer to a remote sampling head, thereby significantly enhancing the versatility of the grazing-angle technique for *in situ* analysis. The development of mid-IR fibres has been somewhat neglected in comparison with silica-based shorter-wavelength counterparts that are used extensively in the telecommunications industry and whose properties are approaching their theoretical limit.[2] Materials that are suitable for fibres and transparent in the mid-IR region are still being developed and improvements in their properties are on-going. The loss spectra of some materials that have been used to fabricate mid-IR optical fibres are compared in Figure 4.3.[3] The loss spectrum is defined here as the attenuation of the radiation travelling through the fibre, due to intrinsic or extrinsic processes related to the fibre material, reflections and fibre bending. At present, the attenuation and physical properties of the available mid-IR fibres mean that they are only suitable for distances of a few metres.[2]

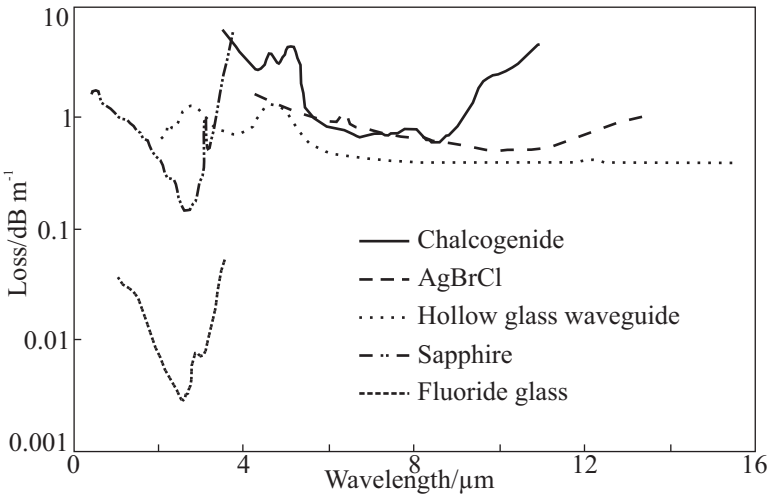


Figure 4.3: Plot reproduced from Harrington,[3] showing the loss spectra for a number of mid-IR fibres.

Chalcogenide glasses are composed of chalcogenides — sulfide, selenide and telluride — chemically combined with elements (usually from group IV or V) such as As, Ge, Sb, P, Ga, Al and Si, to form a glass with two or more components. Common chalogenide fibre compositions are As_2S_3 and As-Se-Te glass; the chalogenide spectrum shown in Figure 4.3 is from a As-Ge-Se-Te fibre. Although the loss spectrum in Figure 4.3 suggests that chalcogenide fibres are not as promising as other materials in terms of there transmission, they have other properties (such as their resistance to moisture and relative flexibility) that make their development favourable.

Chalcogenide fibres are commercially available from a number of sources, including Amorphous Materials Inc. (Garland, Texas, USA) and Oxford Electronics Ltd. (Hampshire, UK). A significant disadvantage is that they are less flexible than their silica counterparts, so fibres, or bundles of fibres, are usually enclosed inside a metal sheath that helps to protect from over flexing, which can result in breakage.

The optical cable used in this research is a bundle of nineteen hexagonal-packed As-Se-Te glass fibres purchased from Amorphous Materials and assembled by Remspec Corporation. Each fibre has a 500-μm diameter core, surrounded by a 50 μm thick As-Se-S cladding, which is coated in thin plastic for mechanical protection. The lower refractive index of the cladding results in total internal reflection of the radiation that is transmitted through the core. The bundle has a total diameter of 3 mm, which is increased to 5 mm by the armoured metal cabling that protects the fibres from over-flexing. The cable is coupled to the spectrometer and sampling head by using Remspec connectors of a similar design to the SMA industry standard. The use of a fibre bundle increases optical throughput while permitting reasonable flexibility. The same throughput through a single fibre would need a larger diameter resulting in unacceptable mechanical properties. The fibres transmit ~90 % over 1 m in the mid-IR, for the range of interest, aside from the H–Se “blind spot” around 2200 cm^{-1}.

4.3 Characterisation of the IR footprint

Characterisation of the shape and intensity distribution of the radiation footprint produced on the surface by the grazing-angle sample head was achieved using a stainless steel T-section reflector (see Figure 4.4) with an arm of 100 mm length and 15 mm width. The grazing-angle sample head was elevated above the working surface until the MCT detector signal had vanished. The cross of the T-section was then placed flush on the underside of the grazing-angle head with the reflective arm perpendicular to the optical direction. The whole T-section was then moved along the 150 mm length of the sampling space and the signal recorded at 10 mm intervals. After several repetitions (to check for reproducibility) the averaged results were graphed to generate a one-dimensional IR profile along the length of the sample-head. To extend this to two dimensions, the reflective arm was covered with a diffusive layer of tissue paper divided into eleven strips, each of 4 mm width, which together spanned the 44-mm width of the sample-head space (see Figure 4.4). With all strips in place, the signal was reduced by ~98 %. The above procedure of intensity measurement was repeated twelve times, removing one strip of tissue each time. The signals were then subtracted from each other to give the signals for each individual strip and the results plotted to give a 2-D profile. This experiment revealed that the IR footprint is elliptical, with principal diameters of ~140 mm and ~30 mm, and that the IR intensity is greatest in the centre, decaying towards the edges (Figure 4.4). This is important since it means that non-uniformity of the sample coating can influence the spectra collected; for example, an area of high concentration in the middle of the IR footprint will give a greater signal than if the concentrated area were in the outer regions.

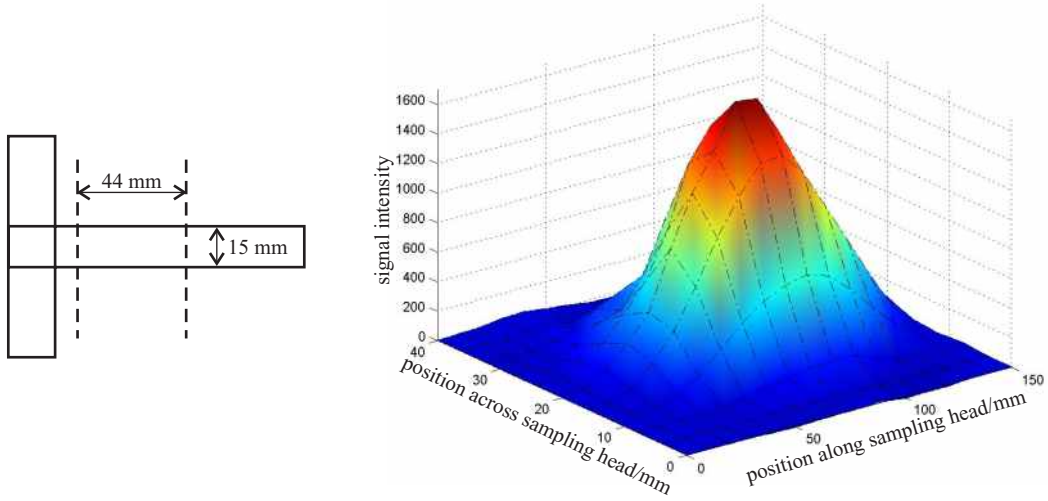


Figure 4.4: Left: schematic of the stainless steel T-section used to determine the IR footprint of the grazing-angle sample-head. Eleven 4-mm wide tissues strips were placed parallel to the dashed lines to span the 44-mm width of grazing-incidence sampling head (indicated by the dashed lines). Right: plot of the 2-D footprint produced by the grazing-angle sample-head.

4.4 Analyte models

The compounds used in this research have been chosen as models for those likely to be of relevance to the pharmaceutical industry. The main compound is the surfactant sodium dodecyl sulphate (SDS: $CH_3(CH_2)_{11}OSO_3Na$, Figure 4.5), which was purchased as a white crystalline powder from BDH Laboratory Supplies, dried at 60°C for two days and then stored in a desiccator prior to use. SDS is an inexpensive detergent and an ionic surfactant that was chosen to be representative of cleaning agents used for industrial equipment. It displays amphiphilic properties where the organic chain forms a hydrophobic tail while the sulfonic head group (OSO_3^-) is hydrophilic.

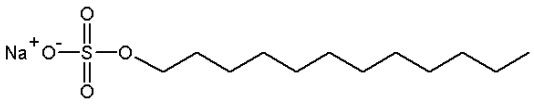


Figure 4.5: Molecular structure of SDS, $CH_3(CH_2)_{11}OSO_3Na$.

SDS is also known as sodium lauryl sulphate (SLS) and both names can be found on the packaging of many household products such as toothpastes, shampoos, shaving foams and bubble baths. It a very effective foaming agent and is used extensively for its thickening effect and ability to create lather. As a detergent, it lowers the surface

tension of aqueous solutions and can be used to emulsify fats. In common with other detergent surfactants, SDS does have some negative properties; it removes natural oils from the skin, which can cause skin irritation, and is also irritating to the eyes.

The infrared absorption spectrum for SDS which has been pressed into a KBr disk is shown in Figure 4.6. The bands of relevance to this work are those around ~3000 cm^{-1} (CH stretches) and in the fingerprint region, ~1500–900 cm^{-1}. The strongest band in the latter region is due to vibrations of the sulfonic head group.[4] The region below 900 cm^{-1} is beyond the range of the MCT detector and has been shown only for completeness.

The IRRA spectra of SDS films are similar to the KBr spectrum in Figure 4.6, the main difference being the strong absorption band of the IR fibre optic at ~2200 cm^{-1}, in an area free from SDS bands. The choice of spectral regions to be used in the analysis of the IRRA spectra is influenced by the presence of interfering compounds. Of particular importance are atmospheric gases such as CO_2 and H_2O, whose effective concentrations can be difficult to control and for which it is easier to avoid regions of absorption. For some analytes, this may not be possible and spectral subtraction and multivariate analysis methods are required.

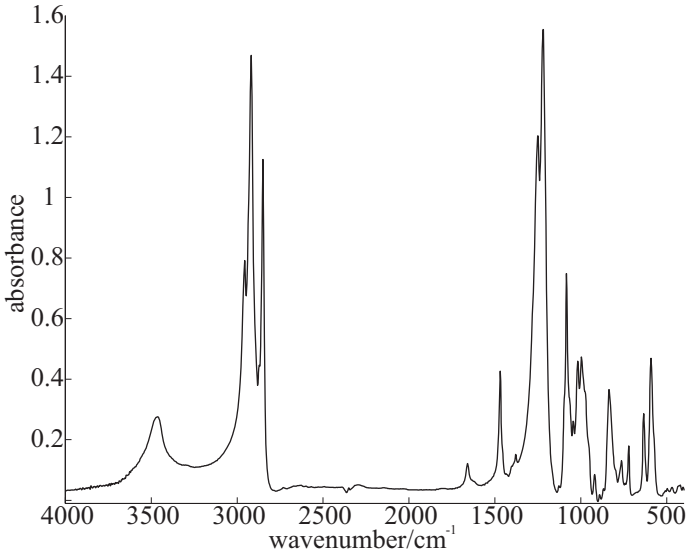


Figure 4.6: Infrared absorption spectrum for SDS incorporated into a KBr disk. The spectrum was measured by the Vector 22 FTIR instrument using the internal sample compartment and DTGS detector.

The second compound is N-acetyl-p-aminophenol (Figure 4.7.), more commonly known by the names paracetamol and acetaminophen, which was purchased from Sigma Aldrich with a stated purity of 98 %. This compound was chosen to be representative of active pharmaceutical ingredients (APIs) and will be referred to hereafter as paracetamol. It is a common analgesic and antipyretic drug that is used for the relief of fever, headaches and other minor aches and pains. It is a major ingredient in numerous cold and flu medications and many prescription analgesics.

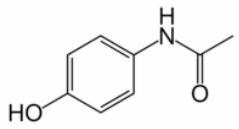


Figure 4.7: Molecular structure of paracetamol, $C_8H_9NO_2$.

The infrared absorption spectrum for paracetamol incorporated into a KBr disk is shown in Figure 4.8, where the 900–400 cm^{-1} wavenumber range has been included for completeness.

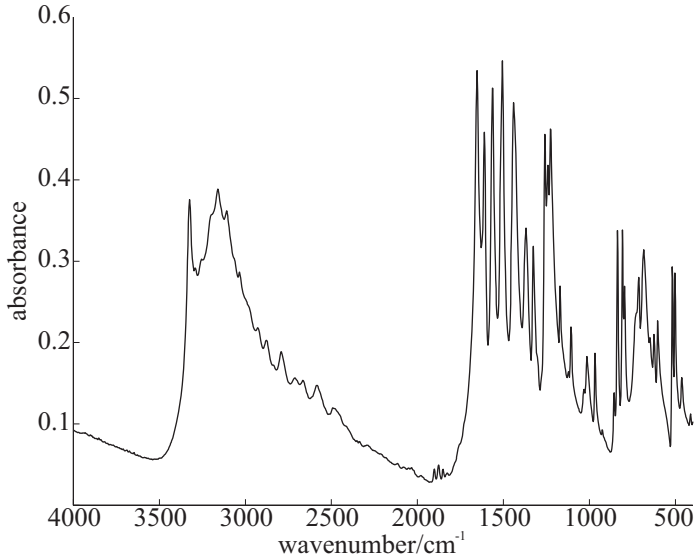


Figure 4.8: Infrared absorption spectrum for paracetamol incorporated into a KBr disk. The spectrum was measured by the Vector 22 FTIR instrument using the internal sample compartment and DTGS detector.

SDS and paracetamol have different physico-chemical properties. For example, SDS is very soluble in water but only slightly soluble in ethanol and acetone, whereas, paracetamol is very soluble in both ethanol and acetone but only slightly soluble in water (although solubility can be increased by heating). The only common solvent in

which both are reasonably soluble is methanol. These solubility differences have led to variations in sample preparation. Other differences have also led to differences in the methods by which the average surface loadings were determined.

For the purpose of comparison with SDS, an industrial cleaner was obtained: P3 cosa® PUR80, manufactured by the German company EcoLab (the P3 cleaner range was originally developed by Henkel Technologies). It is described by the industrial suppliers as "… a neutral, sprayable, concentrated surfactant … suitable for the removal of fatty residues in the pharmaceutical and cosmetic industries."[5] "Excellent removal of organic soiling, such as fats present in residues of creams, ointments, emulsions, Carbopol and tablet coatings", is claimed and also that P3 cosa® PUR80 "enhances the removal of stubborn soiling and is suitable for manual use." [6]

Unlike SDS and paracetamol, which are both white crystalline solids, P3 cosa® PUR80 is a pale yellow liquid with a viscosity similar to that of household detergent. A sample for measurement of the absorption IR spectrum was prepared by smearing a thin film of this cleaning agent over the surface of a blank KBr pellet. The resulting spectrum is shown in Figure 4.9. Not unexpectedly (since it is designated for use as an aqueous cleaner) the sample contains a significant amount of water, as indicated by the broad OH stretching band around ~3500 cm^{-1}. Although the composition of the P3 cosa® is not provided by the manufacturers, Figure 4.9 suggests the presence of long chain organic constituents giving the CH stretching bands around 2900 cm^{-1}. Issues arising from a liquid analyte with unknown composition are discussed in Chapter nine.

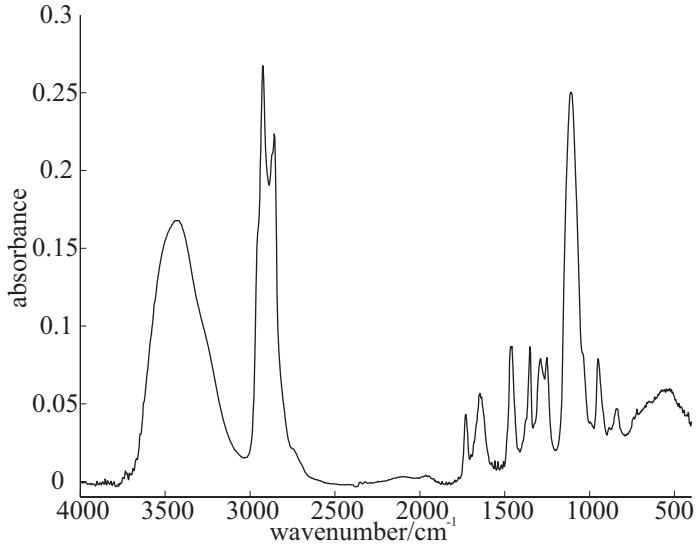


Figure 4.9: Infrared absorption spectrum for the industrial cleaner, P3 cosa® PUR80, from a thin film smeared over a KBr disk. The spectrum was measured by the Vector 22 FTIR instrument using the internal sample compartment and DTGS detector.

4.5 Sample substrates

The sample substrates have also been selected for relevance to the pharmaceutical industry. A range of materials is used, from metals and glass which are used to construct and line industrial vats, to polymer materials used in seals and in tubing to transport solutions around the plant. As discussed in Chapter two, these materials can be classified as metallic or dielectric, with the latter exhibiting infrared features that complicate the IRRA spectra.

The metallic surfaces investigated were aluminium and stainless steel. Of these, stainless steel is used more extensively in the pharmaceutical industry since it is harder than aluminium and has more appropriate mechanical properties. For both substrates, 150 mm × 150 mm plates were cut from larger sheets then cleaned to remove any traces of adhesives, oils or greases. Mechanical wiping was avoided due to the risk of producing microscratches and altering the surface finish. Plates were handled with latex gloves and cleaned by gently wiping with undiluted commercial detergent before rinsing with distilled water. They were then ultrasonicated for 10 minutes in petroleum ether followed by a further 10 minutes with pure ethanol before final rinsing with Milli-Q water. It was important for all substrates that the

detergent be completely removed since it is otherwise likely to interfere with the analysis.

After cleaning, the plates were stored horizontally in racks (to avoid scratching) and rinsed with Milli-Q water before use. To characterise the surface finishes, representative plates were imaged at ×1000 magnification, with the sample tilted at ~45° to the normal, by using a Leica S440 scanning electron microscope (SEM) operating at a beam voltage of 10 kV.

Aluminium was supplied by Wakefield Metal Co. (New Zealand) as 1220 mm × 2400 mm sheets covered with a protective plastic coating. These were cut into plates with care taken to ensure that the sample surfaces remained as scratch-free as possible. The images of cleaned plates (Figure 4.10a) showed that the surface is quite porous. White crystallites, with dimensions ~5-10 μm, seen on the surface are titanium oxide or iron oxide, confirmed by x-ray fluorescence spectroscopy, possibly from polishing processes performed during manufacture of the sheets.

Three different finishes of ~0.73 mm thick 316 stainless steel sheet were purchased from McMaster-Carr (USA). These finishes are referred to in this thesis as “polished”, “smooth” and “rough”. The polished stainless steel has a mirror-like finish, although close observation reveals a faint directional grain due to the polishing process. The smooth steel has the standard mill finish and the rough material has a scoured appearance with no preferential grain. The SEM images of these surfaces are shown in Figure 4.10b-d. They have fewer pores than the aluminium and there are no crystalline impurities on the surface. The differences between the three finishes are clearly evident. The polished finish shows very little detail other than shallow directional grooves from the polishing process. The smooth and rough finishes show far more complexity, with the smooth surface exhibiting a “pancake” like appearance due to the hot-rolling process used to form the sheets. The process used to prepare the rough finish has left deep, larger-scale scores with an overall circular patterning, which is more evident in Figures 6.2 and 6.10 included in the SDS on stainless steel chapter.

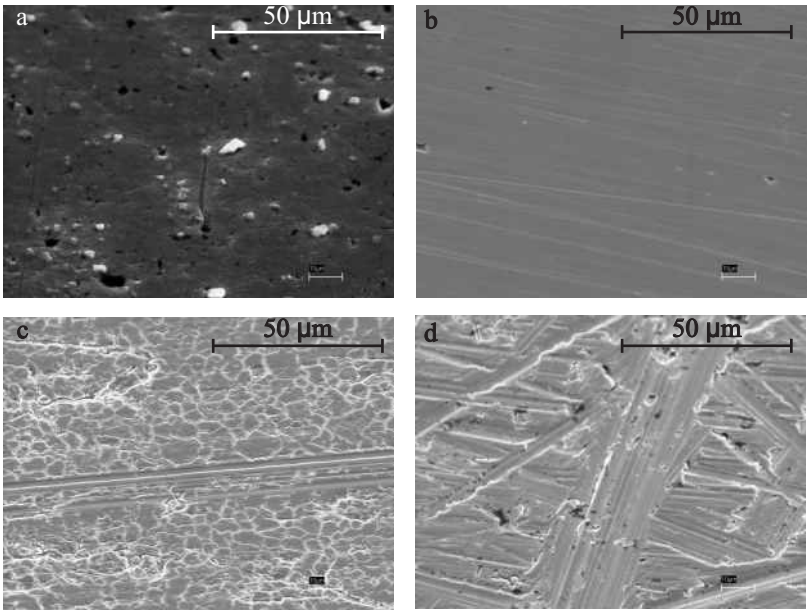


Figure 4.10: SEM images for clean substrate surfaces with a beam voltage of 10 kV. The beam currents were varied according to the nature of the surface and were not recorded for these images. The samples have been tilted 45° to the normal and with a magnification of ×1000. a) aluminium b) polished stainless steel c) smooth stainless steel and d) rough stainless steel.

The dielectric substrates glass, silicone and ethylene propylene-diene monomer (EPDM), were investigated.

Glass plates measuring 160 mm × 160 mm were cut from 3 mm thick window glass. To prevent back-surface reflections, one side of each plate was roughened by bead-blasting using Potters ballotini impact spheres (90–150 µm diameter quartz beads with a density 2.5 g cm^{-3}). This meant that, from the perspective of the IR radiation, the glass plates appear to have an infinite thickness. It was not possible to collect SEM images for the glass substrate. Since glass is harder and less prone to scratching than stainless steel or aluminium, the surfaces were cleaned mechanically by rubbing with solvent soaked paper tissues before being washed with detergent, petroleum ether, ethanol and, finally, Milli-Q water.

In metal or glass-lined vessels, there are likely to be seals and hoses made from polymers and there has been some speculation as to whether the IRRAS technique could be used to quantify contamination on these surfaces.[7] To investigate this, silicone and EPDM (Figure 4.11), purchased from McMaster-Carr (USA), were used as model substrates. These sheets were much smaller than those used for the metal and glass experiments, and were of uneven shape. To obtain consistent results,

50 mm × 140 mm plates were cut to allow them to be accommodated entirely within the IR footprint of the grazing-angle sampling head.

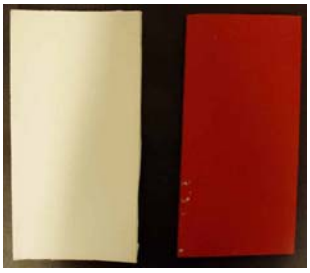


Figure 4.11: Photographs of the EPDM (left) and silicone (right) samples received from McMaster-Carr.

These two polymeric materials have different properties. EPDM (left of Figure 4.11) is slightly stiffer and the corners and edges of sample plates had a tendency to curl slightly upwards. This was alleviated by flattening the substrate with the IRRA sampling head while the spectrum was collected. The silicone surface has an oily texture, while the EPDM surface has a dry texture; both substrates also have rubbery odours. The specifications for EPDM, which is a non-polar polymer rubber, indicate that it has better resistance to polar solvents and can be used in contact with strongly acid and caustic solutions; but it is less resistant to non-polar solvents. Silicone is slightly less resistant to polar solvents but more resistant to non-polar solvents.

The silicone sheets came with white printed type on both surfaces (some can be seen on the left edge of the silicone plate shown in Figure 4.11), which was removed with ethanol. Apart from that, the polymer coupons were ultrasonicated for 10 minutes in each of Milli-Q water and ethanol. It was found, by using UV-visible spectroscopy, that plasticisers from both materials leached into the washing solvent. This would be a concern for pharmaceutical and food industries, but is also important for the IRRAS experiments since the leached materials could constitute an interferent that would be especially problematic if its surface concentration changed over time and/or was dependent on the method by which the sample surfaces were prepared.

Later chapters are dedicated to the results and discussion for each of the substrate surfaces, except for EPDM and silicone, which are covered in a single chapter.

4.6 Sample preparation

To achieve a reliable chemometric calibration the surface loading of the analyte should be as uniform as possible. Only the average surface loading (mass per unit area) of each sample is known and, if a number of spectra are collected from a single sample they are all assumed to have that loading. If there is heterogeneity in terms of coverage distribution and film thickness on length scale approaching that of the infrared footprint dimensions, it will lead to significant uncertainty in the analysis.

Two types of techniques of sample preparation were investigated; spray-based and smear methods. In the spray-based methods, an unknown amount of analyte is applied to the substrate surface and the loading is determined by an independent quantification technique after the IRRA measurements. In the smear methods, a known amount of analyte is applied to the surface, therefore negating the need for independent quantification.

To examine loading uniformity and determine the best method for preparing a uniform film, a visualisation method was required. At the low loadings being investigated it is very difficult to directly image the surface layer in visible light, so a fluorescence technique was employed, using a substitute compound. Anthracene was selected because it fluoresces efficiently in the blue under UV illumination and is soluble in similar solvents to the compounds of interest. Solutions of anthracene in methanol and acetone were used to prepare samples, which were visualised by illumination with UV light in a darkened room. Photographs were taken through a 380 nm low-pass filter, using an Olympus μ300 digital camera. The filter results in sharper images, largely free from interference from UV excitation radiation scattered from the surface. Photographs of clean, unloaded plates showed that all of the observed fluorescence originated from the anthracene (rather than from the plates themselves).

4.7 Smear sample preparation methods

The smear methods are rapid and easily executed without specialized equipment. A known volume (usually 1 mL) of a standard solution with a known concentration of the analyte of interest is dispensed onto the plate surface then spread as evenly as possible, using the straight edge of a soft plastic spatula while the solvent evaporates. The average loading is controlled by altering either the volume or the concentration of the solution. However, the average loading can be subject to error due to the residue that sticks to the spatula, and the point-to-point loading can be uneven due to beading and puddling of the solution.

The uniformity of the surface loading was influenced by the solvent used and the temperature of the substrate. Quicker drying permitted less time for pooling. Slower drying allowed more time to spread the analyte across the surface but also allowed more time for pooling and greater loss of analyte onto the spatula. Other solvent properties, such as volatility and surface tension, influenced the ease with which a uniform surface loading could be achieved.

Smear techniques were trialled on glass, aluminium and stainless steel plates. In the initial investigations, 1 mL of an acetone or methanol solution containing anthracene was placed around the edge of a clean plate, then smeared over the surface until all the solvent had evaporated. The glass and metallic surfaces gave very similar results, as shown in Figure 4.12. A less homogenous finish was obtained with acetone. This can be explained by the greater volatility of acetone, which evaporates too fast to allow the solution to be spread evenly over the entire surface. Methanol permits a more homogenous finish, but still gives a streaky distribution because surface tension causes the droplets to pool.

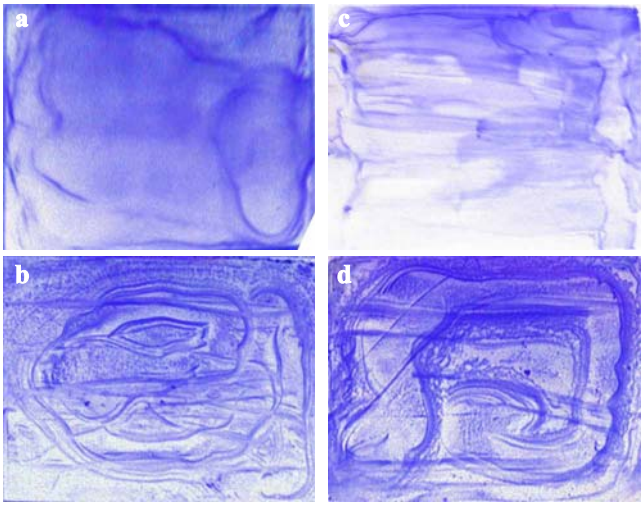


Figure 4.12: Fluorescence images of plates prepared by the basic smear method using 1 mL of a solution of anthracene: a) in acetone on glass; b) in methanol on glass; c) in acetone on smooth stainless steel; d) in methanol on smooth stainless steel.

A number of variations were investigated to improve the performance of the smear method; these included increasing the spatula size, heating the substrate from either above or below to increase the evaporation rate, vibrating the plate and using lipped plates on to which the solution could be poured and left to dry. The results from some of these variations are shown for anthracene on glass in Figure 4.13. Similar results were obtained on metallic substrates. Some of these modifications did result in minor improvements, with methanol solutions tending to give better results, but the surfaces still had streaky and patchy appearances.

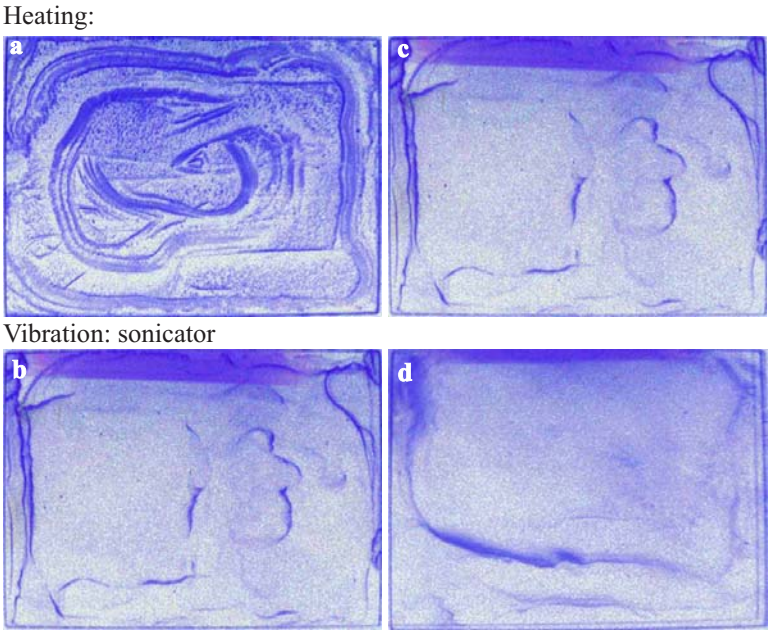


Figure 4.13: Fluorescence images of plates prepared by modified smear methods using 1 mL solutions of anthracene on glass plates. Top row: using a heated (from below) plate, anthracene dissolved in a) acetone and c) methanol. Bottom row: the plate is vibrated by an ultrasonicator during drying, anthracene dissolved in b) acetone and d) methanol.

There are other concerns about smear-based techniques, one of which relates to surface tension. SDS, the compound of main interest in this research, is very soluble in water, but only partially soluble in less polar solvents, such as methanol, and essentially insoluble in acetone and ethanol. Purified water has other significant advantages in being very cheap and easily accessible. But the high polarity, and therefore high surface tension, of water makes it difficult to spread over the surface, so the smear method does not work well with aqueous solutions. This problem is exacerbated by the slow evaporation of water. SDS is a surfactant and should reduce the surface tension. But since anthracene is insoluble in water, it is difficult to visualise surface samples prepared from aqueous solutions. Confidence in the uniformity and consistency of a particular technique is important; and the basic smear technique did not meet these requirements while the modified methods did not result in any significant improvements.

4.8 Spray sample preparation methods

The second investigated type of sample preparation method was spraying. This method involved generation of an aerosol from a solution containing the analyte dissolved in an appropriate solvent.

The first type of spray technique used a Paasche double-action, internal-mix airbrush, operated with a small compressor and ballast tank. It was important that the spray nozzle was not too close to the substrate surface, otherwise excess material caused solvent to stream off the surface causing a lack of uniformity in the loading. Generally, greater air brush-to-surface separation gave more uniform samples, but solvent volatility also needed to be considered. For very volatile solvents, such as acetone, the distance needed to be no more than 15–20 cm to prevent significant droplet loss due to evaporation. For ethanol or methanol, the distance from the plate could be increased ~40 cm, and with water, it was increased to ~80 cm. For the shorter spraying distances, the air brush had to be swept carefully and evenly across the surface to achieve the most homogeneous loadings.

The surface loading could be varied by changing the concentration of the solution or the amount of solution sprayed on the substrate; the latter was dependent on the number of passes of the airbrush and the total spraying time. Samples could be loaded with more than one compound either by spraying separate solutions consecutively or by spraying a mixed solution.

Fluorescence visualisations of glass and polished stainless steel plates sprayed with a solution of anthracene dissolved in methanol are shown in Figure 4.14. It is clear that the spray method produces samples that are far superior to those prepared by the smear method. The coverage stretches to the corners of the sample surface, and tests showed the apparent concentration gradients are almost entirely due to the positioning of the UV light source and the image capture of the camera, (The purple band across the top of the stainless steel image is a reflection of the UV lamp).

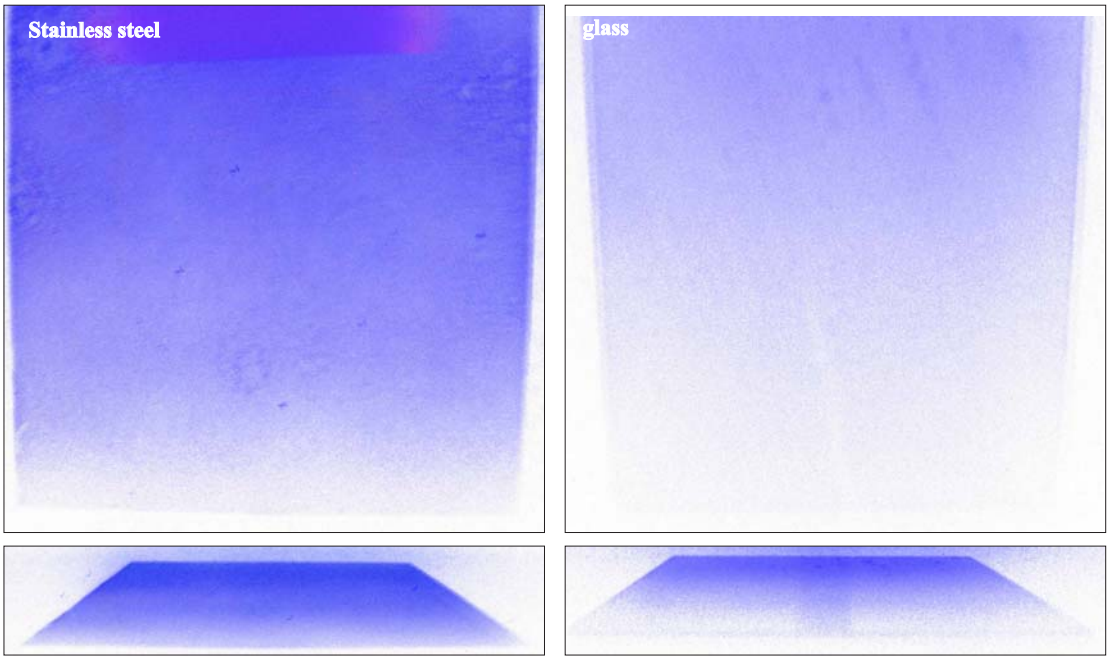


Figure 4.14: Fluorescence visualisations from anthracene sprayed onto polished stainless steel (left) and glass (right) surfaces using methanol as the carrier solvent. The glass surface has the smaller surface loading. The lower images are photographs of the same plates taken at oblique angles.

Another spray based method employed a medical nebuliser of the type used for inhalation administration of drugs, for example in the treatment of asthma. This method produces very fine aerosol particles (mean diameter 5 μm) that evaporate very rapidly. In trials using in methanol and water solutions, it was very difficult to obtain any detectable loadings on the surface, even when the nebuliser nozzle was held very close, <1 cm to the plate, since the solvent evaporated too rapidly and the dry analyte particles would not adhere to the surface. Several modifications were attempted to overcome this problem. In case the aerosol droplets were electrically charged, the aerosol flow was passed between two oppositely charged (±50 V) metallic plates, but the loadings (as judged by IRRAS measurements) were still negligible.

Apart from a few cases where the smear method was more appropriate, spray based methods using the Paasche air brush were used in experiments described in this thesis. Figure 4.15 shows SEM images for samples of SDS at ~1 μg cm^{-2} and paracetamol at ~2 μg cm^{-2} prepared by this method on polished stainless steel. These show that samples prepared by the spray method are not thin films, but rather collections of spots formed in the spraying process. These images also indicate that average spot thicknesses are different for different analytes; *viz* of the order 10 nm for SDS on

polished stainless steel and 100 nm for paracetamol on polished stainless steel on the left and right respectively of Figure 4.14.

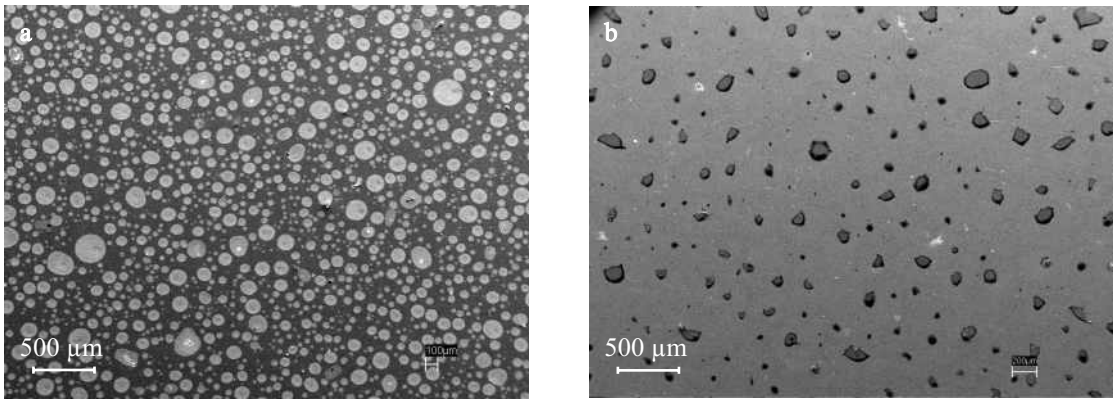


Figure 4.15: Low-resolution (×90) SEM images of SDS at a loading of ~1 µg cm^{-2} (left) and paracetamol at a loading of ~2 µg cm^{-2} (right) on polished stainless steel. The beam voltage was 10 kV for both images; the beam current was 3 nA for the SDS and 100 pA for the paracetamol. The images are taken looking down the surface normal at working distances of ~8 mm for SDS and ~10 mm for paracetamol.

4.9 Primary calibration methods

Since the amount of material deposited by the spray method is difficult to predetermine, an independent method is needed to establish the true loading for correlation with the IRRA spectra. Attempts to use differential weight measurements (using a 10 µg resolution Mettler AE101) relied on small changes (mg) of much large (~100 g) values; these were not reproducible within the accuracy required and were abandoned. The quantitative dissolution of the analyte from the plate followed by a spectroscopic quantification proved to be much more reliable. A minimum volume of 50 mL of an appropriate solvent was required to quantitatively remove the analyte, but use of 100 mL provided a margin of safety. Water or methanol was used for SDS and ethanol for paracetamol.

Although a single method for SDS and paracetamol quantification would have been preferred it quickly became clear that this was impractical and that different calibration methods would be required. Since paracetamol has a strong UV absorption band ($\varepsilon_{249} = 13700$ L mol^{-1} cm^{-1}), univariate UV colorimetry proved to be a convenient, sensitive and accurate method for determining the surface loading of that

analyte. A GBC-920 UV/visible spectrophotometer was used at 2 nm resolution and a scanning speed of 180 nm min^{-1} over the wavelength range 210–340 nm; the relevant spectrum is shown in Figure 4.16.

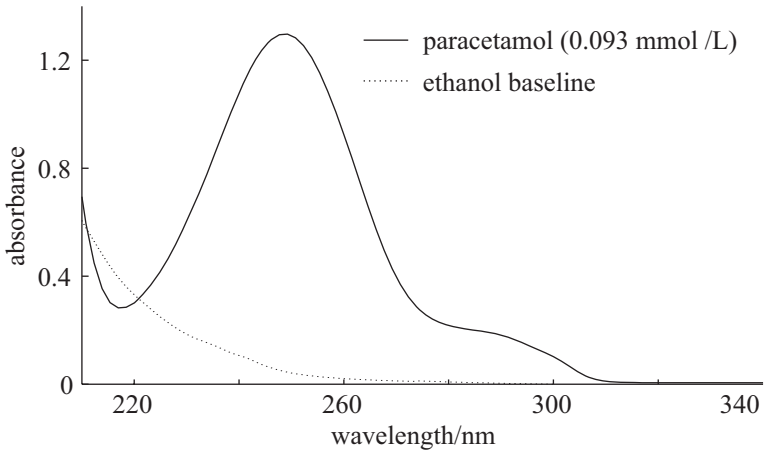


Figure 4.16: UV/visible spectrum of paracetamol at a concentration of 0.093 mmol L^{-1} (14 mg L^{-1}) in ethanol over a 1 cm pathlength. The background spectrum of ethanol is shown by the dashed curve.

A calibration curve prepared by measuring the absorbance at 243 nm for a series of freshly prepared standard solutions of paracetamol in ethanol is shown in Figure 4.17. The washings concentration for each of the unknown samples was determined from such calibrations and the average surface loadings were back-calculated. UV/visible colorimetry is a very sensitive technique and the absorbance at 243 nm resulted in a very good calibration, hence the application of multivariate calibration methods was not considered necessary.

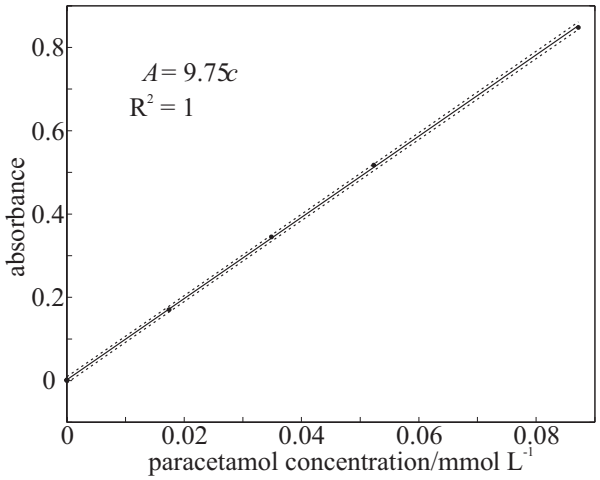


Figure 4.17: UV calibration for paracetamol absorbance at 243 nm with the 95 % prediction interval shown as dashed curves. The calculation of the linear regression calibration and associated statistics are covered in Appendix A.

Since SDS does not absorb in accessible regions of the UV or visible radiation ranges, an alternative method was required for this analyte. HPLC was ruled out since the UV and refractive index detectors of the available instruments gave unsatisfactory results.

Considerable effort was made into investigating atomic absorption (AA) spectroscopy to determine the amount of sodium, and hence the amount of SDS (assuming a 1:1 molar relationship), present in the washings. The instrument used was a Varian SpectrAA 220 FS operated with Spectra AA version 4.10 PRO software. The measurement wavelength was 589.0 nm and the instrument slit width and lamp current were 0.5 nm and 5.0 mA, respectively. The flame was produced by acetylene (flow rate 2 L min^{-1}) with air support (13 L min^{-1}). A calibration was prepared from the change in signal intensity between blank and standard solutions.

There were a number of problems with this method, the most serious being the ubiquitous levels of environmental sodium, which required great care to avoid contamination. To avoid the foaming problems with water, methanol was chosen as the solvent, but it was subsequently found that solvent grade methanol contains a reasonable amount of Na, (analytical grade methanol has less Na content but is prohibitively expensive in the volumes required in this research). Double distillation reduced the Na to an acceptable level, but other issues arose with some of the Na from the SDS going undetected due to processes such as self absorption and ionisation. CsCl was added as an ionisation suppressant.

Standards were used for the calibration and were also measured along with samples from the plate washings to assess experimental drift and detector scatter. Due to slow leaching of Na from glassware, fresh standards were prepared for each set of unknown samples. These precautions revealed that the AA method suffered from poor reproducibility, mostly associated with the instrument. Figure 4.18 shows results from different batches of standards run under nominally the same instrumental conditions. Batch one results were those obtained initially; batch two after replacement of the Na lamp; and batch three after adjustments of the sample injection system. These reproducibility issues and the risk of Na contamination meant that the AA method could not cope with low-concentration samples and it was eventually abandoned.

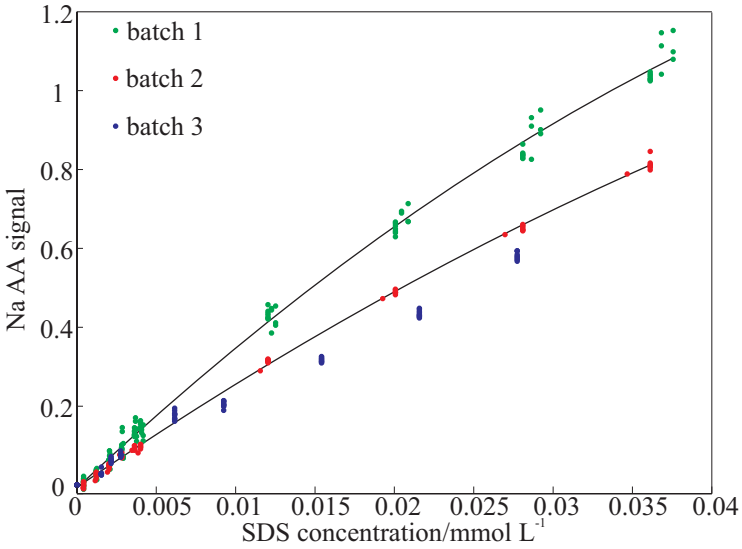


Figure 4.18: Calibration curves for quantification of SDS by Na atomic absorption; the batches refer to experiments run while the instrumental conditions were nominally constant – see text.

^{1}H NMR spectroscopy was found to give the best combination of sensitivity and reproducibility for SDS quantification. The SDS was rinsed from the plates using Milli-Q water. The required solvent volumes precluded the use of deuterated solvents, so suppression techniques were applied to the NMR spectra to prevent the water signal (at 4.79 ppm relative to tetramethyl silane (TMS)) from swamping the analyte signal. Water was used rather than methanol since it has only one ^{1}H NMR resonance (rather than two) and hence requires suppression at only one frequency. Foaming was avoided by carefully inverting the washings several times before they were analysed. The ^{1}H NMR spectrum for SDS in D_2O (Figure 4.19) shows four distinct groups of resonances, the most intense of which (~1.2 ppm) is attributed to the protons in the CH_2 backbone.

Quantitative ^{1}H NMR, requires referencing against an internal standard that has a simple spectrum that does not interfere with the spectrum of the analyte. A single resonance in the vicinity of, but not overlapping with, the analyte signal would be ideal in order to avoid phasing and baseline distortions (which are commonly introduced by solvent suppression). A large separation can also influence the intensity ratios used for the calibration. Another important consideration is the need for the internal standard signal to be well separated from the suppressed solvent peak. From inspection of Figure 4.19 a good internal standard would have a signal in the region

1.6–3.8 ppm. Of several possibilities, dioxane, with a singlet at 3.75 ppm, was selected because of its low volatility and because it was less likely than other candidates (such as methanol) to be influenced by contamination. A 500 μmol L^{-1} solution of dioxane in a phosphate buffer at pH 7.5 was found to very stable and did not need to be prepared freshly for every set of experiments.

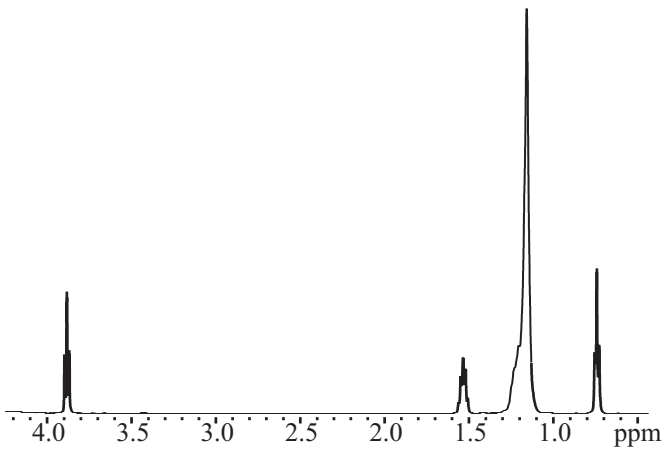


Figure 4.19: ^{1}H NMR spectrum for SDS dissolved in D_2O.

Samples prepared by adding 100 μL of buffered dioxane solution to 400 μL of SDS rinsate were placed in a 5 mm NMR tube with a 3 mm D_2O insert. An Oxford Instruments AS500 spectrometer was used to collect thirty-two transients in 9 minutes, using 8.1 μs, 90° radio frequency pulses and 1.89 second acquisition time, initially with a 15 second delay between pulses. The long delay time was used to permit excited molecules to relax between pulses, was later reduced to 5 seconds without loss of sensitivity but permitting thirty-six transients to be collected in just 4 min. Each transient comprised 30,272 data over a sweep width of 8 kHz. The 4.79 ppm H_2O proton singlet was suppressed by using a double-pulse field-gradient spin-echo (DPFGSE) technique. The transients were zero filled to 128 k data points prior to Fourier transformation and any phasing or baseline corrections were carried out manually by using Varian VNMR software, version 6.1 C or D. A spectrum for a 0.0139 mmol L^{-1} SDS standard is shown in Figure 4.20.

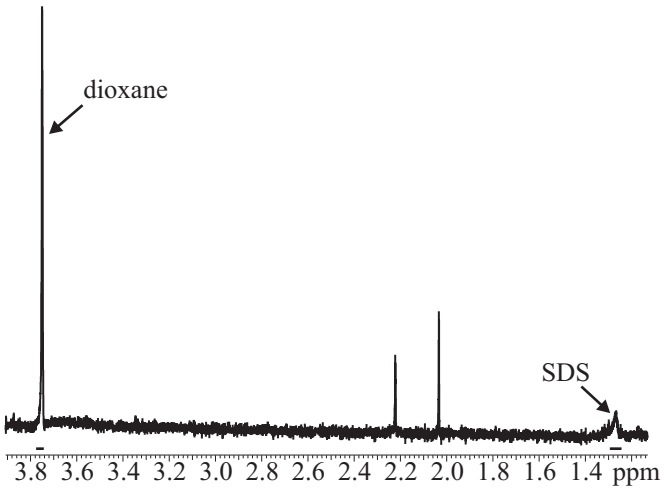


Figure 4.20: ^{1}H NMR spectrum for a 0.0139 mmolL^{-1} SDS standard, the integral regions for the dioxane and SDS peaks are shown as a horizontal bar above the axis. The two unlabelled peaks are contamination peaks.

For each rinsate, several NMR spectra were measured to ensure reproducibility. The integrals of unresolved SDS CH_2 resonances (1.294–1.244 ppm) and the dioxane resonance (3.757–3.744 ppm) were determined over the ranges indicated by the horizontal bars above the abscissa in Figure 4.20. The ratio of the SDS to dioxane integral was then used to produce a linear calibration over the SDS concentration range 0–0.084 mmol L^{-1} (Figure 4.21). The range was reduced over the course of the research as smaller surface loadings became more important. In order to ensure that the prepared NMR calibration remained stable a standard solution was included with each set of NMR samples.

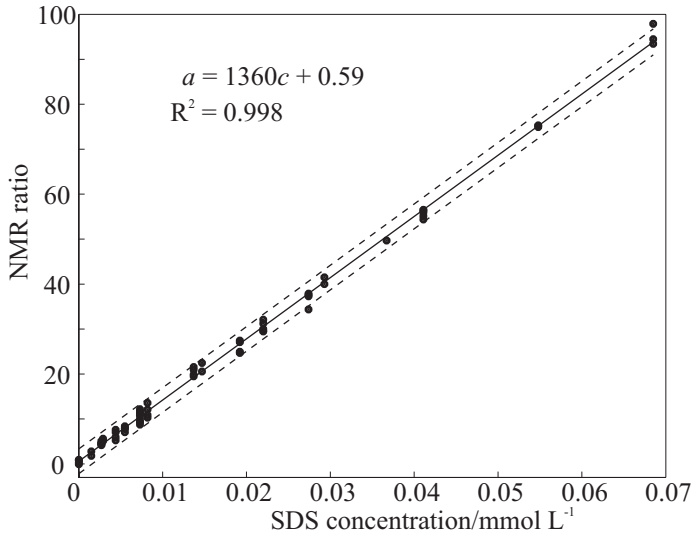


Figure 4.21: Typical ^{1}H NMR calibration for SDS. The 95 % prediction interval is shown by the dashed lines. The procedure for calculating a univariate prediction interval is described in Appendix A.

A number of irritating contamination issues had to be overcome before the NMR calibration was adopted as the primary calibration method for SDS. The two unlabelled peaks in Figure 4.20 appeared in all NMR spectra collected with a particular D_2O insert, and are probably due to contamination of the D_2O by acetone and another unknown compound. The insert was subsequently replaced without affecting the calibration. An unidentified contaminant produced an NMR triplet of variable intensity that overlapped with the left hand side of the ~1.2 ppm SDS peak. Attempts to identify and eliminate the contaminant were unsuccessful, but it was clearly related to the D_2O insert and could be minimised by keeping the insert very clean. To ensure that it would not interfere with the quantification of unknowns, the SDS integral range was reduced without adverse effects on the calibration. Another problematic contaminant was ethanol, which has a triplet at 1.17 ppm and a quartet at 3.65 ppm and interferes with both dioxane and SDS. This solvent was commonly used by other laboratory users, and even small amounts of it in the atmosphere were enough to interfere with the NMR experiments. The only solution to this problem was to perform these experiments on days when ethanol was not to be used or to conduct the experiments early in the mornings before the other workers were around!

4.10 References

1. Melling, P. J. and Shelley, P. H. Spectroscopic accessory for examining films and coatings on solid surfaces. US patent no. 6310348, 2001.
2. Lendl, B. and Mizaikoff, B. In *Handbook of Vibrational Spectroscopy: Sampling Techniques*; Chalmers, J. M. and Griffith, P. R.; John Wiley and Sons: Chichester, UK, 2002; Vol. 2.
3. Harrington, J. A. *Infrared Fibers and Their Applications*; SPIE, 2004.
4. CRC *Handbook of Chemistry and Physics*, 56 ed.; CRC Press, Inc: Cleveland, Ohio, 1975.
5. BS International. Ecolab, http://www.bsibusiness.com/ecolab.asp.
6. Ecolab. Ecolab: pharmaceutical / cosmetics, http://www.ecolab.com.
7. Thomson, M. and Melling, P. J.; Hamilton, M. L., 2006, Personal communication regarding: industrial applications of the IRRAS technique.

Chapter 5

SDS on an aluminium substrate

5.1 Introduction

It has been reported in the literature that IRRA spectra of organic films adsorbed on aluminium surfaces correlate well with surface loading.[1, 2] In those studies, the films met the criteria in Chapter two, where the analyte film forms a homogeneous layer over the entire, optically flat, substrate surface. As discussed in Chapter four the samples investigated in the research reported in this thesis do not meet the same criteria. But, since quantitative relationships between IRRA spectral signals and surface loadings of well-behaved films on aluminium have already been established, a study of SDS films on an aluminium substrate provides a good starting point to evaluate whether inhomogeneous samples can be quantitatively evaluated.

5.2 Sample preparation

The aluminium plates and their preparation are described in Chapter four. Samples of SDS on aluminium (SDS:Al) were prepared by using the spray method, with Milli-Q water as the solvent, to give surface loadings in the range 0–2.4 μg cm^{-2}. Eight spectra were collected from each sample, four with the optical path along one axis of the plate and four others in the perpendicular direction. The average surface loadings were determined from aqueous washings using the quantitative ^{1}H NMR method described in Section 4.9.

A typical IRRA spectrum for SDS:Al is shown in Figure 5.1 at a loading of 2.36 μg cm^{-2}, at the upper end of the calibration range. The horizontal bars span the regions that contain the SDS peaks used in the quantitative analyses described below.

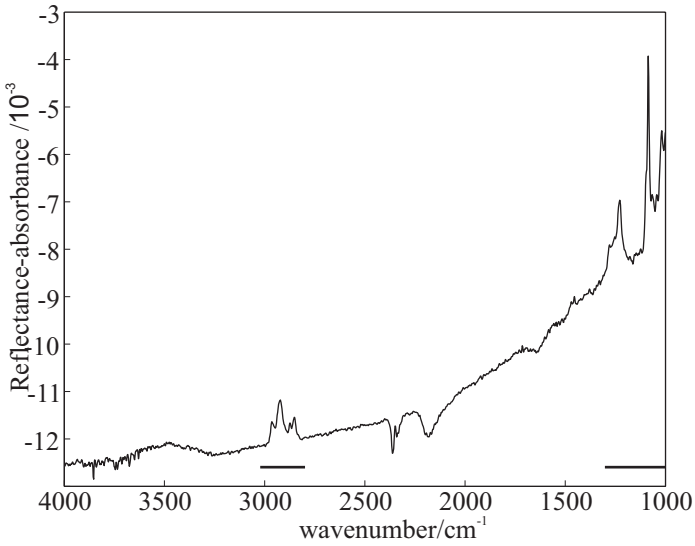


Figure 5.1: IRRAS of SDS on aluminium surface with a loading of 2.36 μg cm^{-2}.

5.3 Quantitative analysis using peak area

Initially, a peak-area method was used to test whether the loadings of the "spotty" SDS samples (see Figure 4.14) could be determined quantitatively using univariate analysis. Integration was carried out using the trapezium method, subtracting a straight line that runs through the first and last points in the integration range. The bands in the fingerprint region (1000–1300 cm^{-1}) were found, empirically, to give much better results than those in the CH stretch region (~2900 cm^{-1}). This is not due to baseline problems, since the fingerprint region is more strongly affected by steeply sloping, non-linear baselines. Contamination by organic compounds is more likely to cause interference in the higher energy region, since the CH stretching bands will overlap strongly making it difficult to select those bands specifically due to SDS. In the fingerprint region, the sulfonic head-group bands are better defined, more greatly separated and less likely to be replicated by common contaminants. In addition, the shapes of the bands in the CH region are known to be influenced by orientation and the nature of the surface adsorption.[3]

Figure 5.2 is a plot of SDS loading (0–2.4 μg cm^{-2}) *versus* peak area for SDS:Al obtained using the integration range 1209–1292 cm^{-1}. The averaged area for the eight spectra from

each sample has been plotted since this gives a better representation of the average loading. Also displayed (as dashed curves) are the 95 % prediction interval limits (see Appendix A). Although the plot is reasonably linear, the prediction interval range of appropriately ± 0.38 μg cm^{-2} corresponds to a DL of ~0.4 μg cm^{-2}. As shown below, better results can be achieved using multivariate analysis.

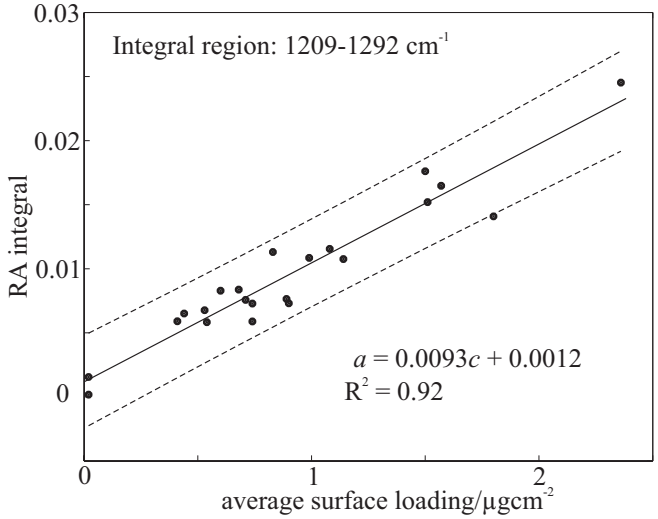


Figure 5.2: SDS surface loading *versus* peak area for the SDS peaks in the 1209–1292 cm^{-1} region of the IRRA spectrum of SDS:Al. The 95 % prediction interval, indicated by the dashed lines, corresponds to a loading uncertainty of ~0.75 μg cm^{-2}.

5.4 Quantitative analysis using PLS

Better calibration models are likely using multivariate statistics, which confer the advantages of resolving issues to do with baseline distortions and interfering spectra from contaminants and atmospheric gases (Chapter three).

PLS models were built for the combined SDS spectral regions shown in Figure 5.1 and also for each region individually. The three data averaging approaches, NA:NA, NA:A and A:A (as described in Section 3.12) were initially applied without data pre-treatment, and the plots of RMSECV against rank are displayed in Figure 5.3 (NA:NA crosses; NA:A circles; A:A triangles). As explained in Section 3.12, the leave-one-spectrum-out, NA:NA, approach makes the analysis susceptible to bias and is included here for completeness. The best calibration models for the two leave-one-sample-out methods

(NA:A and A:A) were selected, as described in Chapter three, to give the lowest rank that has a corresponding REMSECV that is not significantly greater than the minimum value, these are indicated in Figure 5.3 as solid symbols.

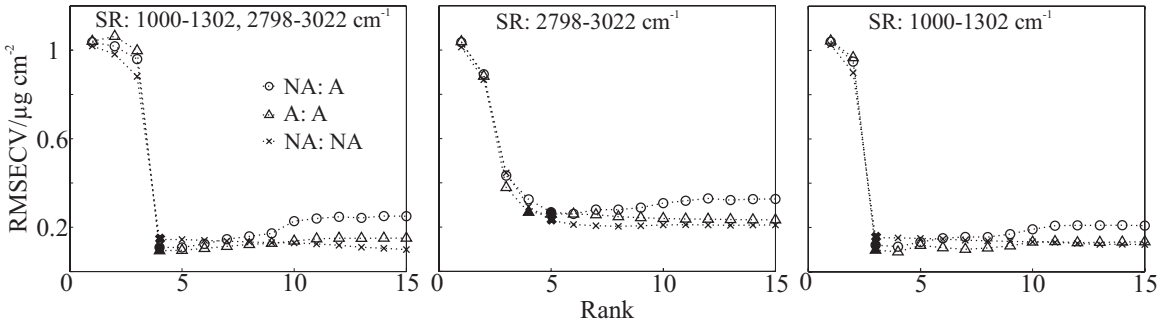


Figure 5.3: RMSECV *versus* rank plots for SDS:Al using the spectral regions (SR) indicated. The NA:A data refer to calibrations where the PLS calibration is built using the non-averaged spectra (i.e. all eight spectra per sample and considered individually), but all spectra from the sample are removed during the cross validation and their predicted loadings are averaged. In A:A series, the averaged spectra have been used in both the calibration and cross validation. The NA:NA series are where all spectra from all samples are treated independently through the entire calibration and cross validation processes. The optimum rank for each system is denoted by the solid symbol.

The performance of the calibration model is very dependent on the uniformity of surface coverage and the accuracy with which the true loadings are known. It was noted in Chapter three that for carefully prepared samples the differences between the two leave-one-sample-out approaches should be very similar. Figure 5.3 shows that this is indeed the case for the PLS models for SDS:Al, independent of spectral ranges that are used, and suggests that the samples were reasonably homogeneous.

The model built using the CH region (2798–3022 cm^{-1}) is inferior to those using the fingerprint or combined regions, with a greater RMSECV minimum of ~0.3 μg cm^{-1} compared with ~0.1 μg cm^{-1}. The model built on the fingerprint region alone (1000–1302 cm^{-1}) has a lower optimum rank (3 compared with 4-5) but a similar RMSECV to that of the combined model. The overall conclusion, that the fingerprint region gives the best model, is consistent with the results from the univariate (band-area) analysis. The justification regarding the influence of contamination and interference in the CH region (see section 5.3 above) is reinforced by the fact that fewer PLS factors are required to achieve a good calibration when that area is excluded from the analysis.

The predicted *versus* true loading plots for the NA:A data treatment using the optimal ranks in Figure 5.3 are shown in Figure 5.4. The dots represent the values for the non-averaged spectra and the open circles the average results for each sample. The scatter appears to be reasonably homoscedastic, with the distribution of non-averaged data around their averages being fairly independent of loading. It was expected that the heavier surface loadings would display more scatter since uniform films are more difficult to form at higher loadings.

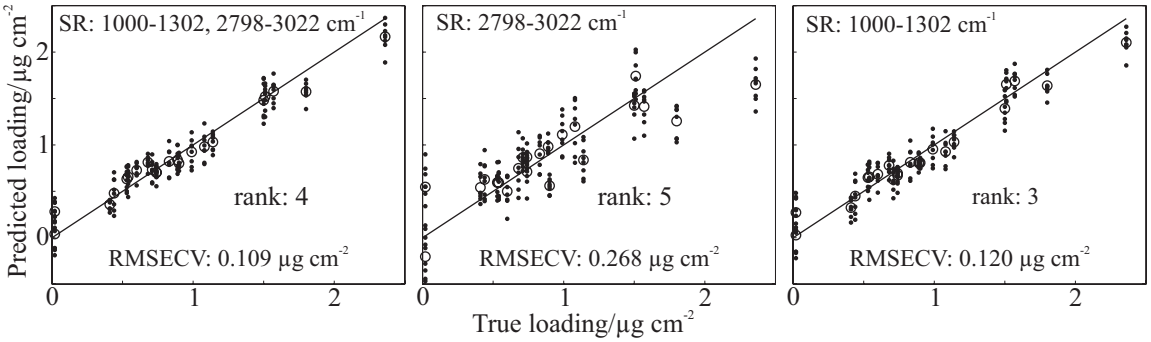


Figure 5.4: The predicted *versus* true loading plots for the NA:A data treatment using the optimal ranks in Figure 5.3. The dots represent the values for the non-averaged spectra and the open circles the averaged results for a single sample.

The possibility of statistical outliers was investigated by comparing the leverage (h) for each spectrum with its loading residual, as described in Section 3.12. Figure 5.5 contains plots of leverage *versus* standardised loading residuals for all spectra included in the final calibration set. Outlier spectra which were removed from the calibration set were detected by analysing both their leverages and their concentration residuals. If a spectrum was found to have both a large residual *and* large leverage then the entire sample (all eight spectra) was removed from the calibration. Spectra that had either a large residual *or* a large leverage were tagged as possibly troublesome. Decisions on whether they were to be retained in the calibration set were made on a spectrum-by-spectrum basis, relating to the surface loading and the values of the leverages and residuals of other spectra from the same sample. In the model using just the CH region, there is one spectrum that has a standardised residual value marginally greater than +3 standard deviations. However this spectrum has a small leverage ($h \approx 0.04$) and was not removed from the calibration set.

The combined- and the fingerprint-region models calibration sets contain no data with unduly large leverages or residuals.

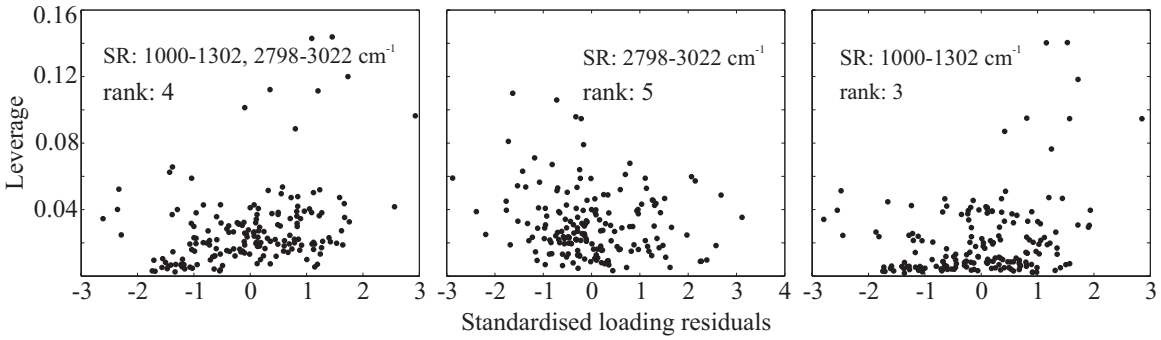


Figure 5.5: The leverage *versus* standardised loading residuals for the spectra in the PLS calibration models shown in Figure 5.4.

5.5 The use of pre-treatments

Several pre-treatments (Section 3.3), implemented using the OPUS software and/or Matlab scripts, were applied to the data and the results were compared to the results without pre-treatment. The pre-treatments included:

- Mean centring, which is commonly applied to either or both the spectral and concentration data matrices to control leverage;
- Constant offset subtraction, by which the spectral minimum is set to zero;
- Straight-line subtraction fits, which correct for a linearly sloping baselines;
- First-derivative smoothing; and
- Second derivative smoothing.

If a pre-treatment works well, the resulting RMSECV *versus* rank plot should show a minimum at a lower rank than is the case without pre-treatment, with either the same or a smaller RMSECV value.

Since the three data-averaging approaches in Figure 5.3 displayed similar behaviour, only one of them was used to investigate the effects of data pre-treatments; and because it was the easiest to which to apply the pre-treatments, the A:A approach was chosen. The PLS models were built using the combined spectral regions, so that all of the spectral information was available to the model and to see whether the pre-treatments could improve the "behaviour" of the CH region. The resulting RMSECV *versus* rank plots are presented in Figure 5.6. The top left plot is the same as shown in Figure 5.3 for the A:A approach applied to the combined spectral regions; it has been included for completeness and ease of comparison. The optimum ranks, as judged by the methods described in Sections 3.8 and 3.12 are indicated by the filled circles on the plots, and the numerical results are summarised in Table 5.1.

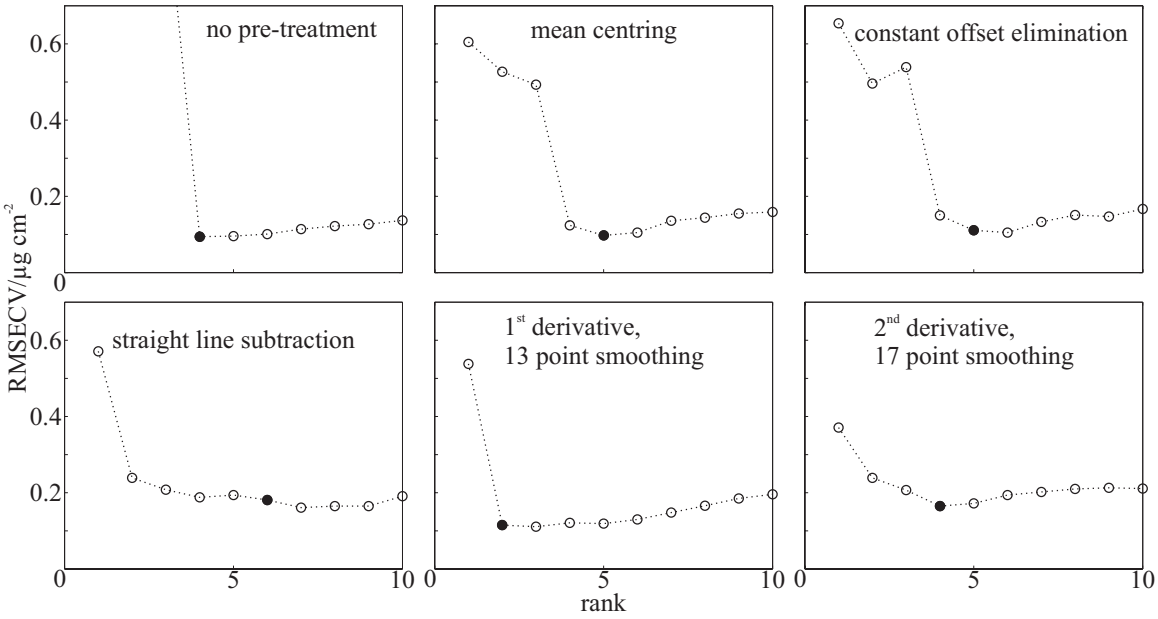


Figure 5.6: The RMSECV *versus* rank plots for different spectral pre-treatments applied to the A:A data treatment for the SDS:Al data for PLS calibrations using both SDS spectral regions. The optimum rank for each model is indicated as a filled in circle.

These results show that none of the pre-treatments offers advantages. All increased the minimum REMSECV (in some cases only slightly) and most gave a larger optimum rank. The only one that appears to offer a partial improvement is 1st derivative smoothing, where optimum rank is reduced from four to two with a RMSECV that is only slightly

higher. Further investigations regarding the amount of smoothing showed that the lowest RMSECV values were achieved when 9–13 point smoothing was carried out.

Table 5.1: PLS parameters for the best PLS calibrations selected in Figure 5.6 for the different spectral pre-treatments.

Pre-treatment	**Smoothing**	**Optimum rank**	**RMSECV /µg cm^{-2}**
none	–	4	0.0943
mean centring	–	5	0.0975
constant offset elimination	–	5	0.111
straight line subtraction	–	6	0.181
1st derivative	13 points	2	0.115
2nd derivative	17 points	4	0.165

5.6 Detection limits

Although the determination of accurate detection limits (DLs) is difficult, the techniques described in Chapter three can be used to estimate these statistics using information from primary calibrations and PLS models. The determination of detection limits, covered in Section 3.12, assumes that leverage is small and that the variance associated with the concentration measurement errors in the reference method is also negligible.

The leverages for each of the samples, shown in Figure 5.5, were calculated using equation (3.71). The simplification of the σ_{PE} expression (equation (3.68)) that is required in the determination of DLs (equation (3.69)) depends on the leverage for all samples in the calibration set being small. The leverage of a spectrum is related to its distance from

the origin of the factor space, and in the absence of mean centring, should increase with surface loading,[4] a trend that is followed by the SDS:Al spectra, as shown in Figure 5.7. But despite this trend, all leverage values in Figure 5.7 are all significantly less than one.

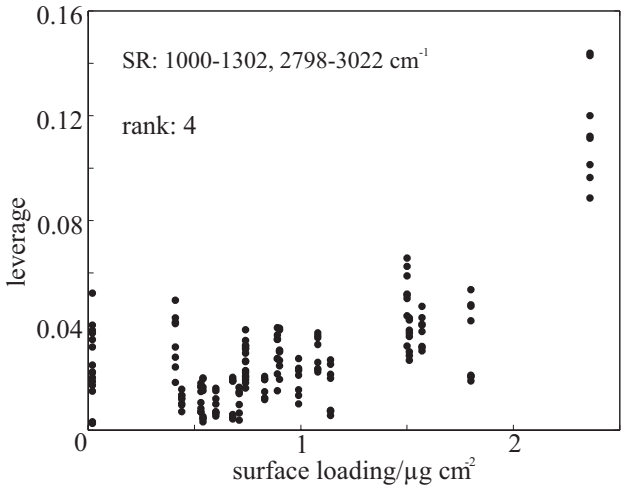


Figure 5.7: Plot of sample leverage for non-averaged spectra during cross validation in the PLS calibration built using the combined spectral regions.

The second assumption, that concentration measurement errors in the reference method are negligible, is known to not be true. There are errors associated with the concentration values determined from the primary NMR calibration but, more importantly, there are also errors introduced by the non-homogeneity of the SDS coverages, as seen in Figure 5.4. This factor can be ignored within the pessimistic approach to errors adopted in this thesis, since the term $\sigma_{\Delta c}$ leads to improvement of the calculated DLs. Another consideration is that the two terms that are ignored, the leverage and concentration errors, work against each other (see equation (3.68)) leading to cancellation.

The final assumption, that RMSECV can be used to model the RMSEP, results in the expression for the DL shown in equation (5.1).

$$\mathrm{DL} = t_{1-\alpha,\rho} \times \mathrm{RMSECV} \qquad (5.1)$$

The 1-α DL (e.g. if $\alpha = 0.05$ it is the 95% detection limit) is calculated using the appropriate *t*-statistic for a system with ρ degrees of freedom (DOF). As discussed in

Section 3.12, there are difficulties in determining the effective DOF. The conservative approach adopted throughout this thesis is that the DOF is approximated by $\rho = m - r$; m being the number of samples (twenty-six in this case) and r being the rank (four using the NA:A data treatment on the combined spectral regions) for the PLS calibration.

Table 5.2: Detection limits for the best PLS models denoted in Figure 5.3 for calibration models built using non-averaged spectra and averaged predictions.

spectral region	wavelength range/cm^{-1}	Optimum rank	ρ	$t_{0.95,\rho}$	RMSECV /μg cm^{-2}	DL /μg cm^{-2}
combined	1000-1302 & 2798-3022	4	18	2.10	0.109	0.23
CH	2798-3022	5	17	2.11	0.268	0.56
fingerprint	1000-1302	3	19	2.09	0.120	0.25

5.7 Conclusion

For samples of SDS on aluminium prepared using the spray technique, there is a quantitative relationship between the surface loadings and the IRRA spectra. This was verified by using both univariate and multivariate statistical techniques. The results achieved using the multivariate PLS technique are superior to the univariate band area results, which is indicated by (pessimistic) DLs for the PLS models that are almost half (~0.23 μg cm^{-2}) that calculated for the univariate calibration (~0.4 μg cm^{-2}). The better performance is attributable to the PLS calibration's ability to discriminate contaminants and baseline effects.

PLS calibrations were generated using the fingerprint and CH stretching regions of the SDS:Al IRRA spectra separately, and also the combined regions. Leave-one-sample-out and leave-one-spectrum-out cross-validations were tried, with two variations (NA:A and

A:A) of the latter. The fingerprint region, using the NA:A approach gave the best results, but those using the combined regions were only slightly inferior with the larger optimal rank but a slightly lower corresponding RMSECV. The models built using the CH region only gave the poorest results.

The effects of spectral pre-treatments on the averaged spectra calibrations were investigated and it was found that the best results were achieved when no pre-treatments were applied. In fact, in many cases the pre-treatment resulted in a deterioration of the calibration, which was unexpected. This led to the decision to exclude pre-treatments of the spectral data in subsequent analysis.

5.8 References

1. Gaillard, F., Linossier, I., Sweeney, M., Reffner, J. A. and Romand, M. Grazing angle micro-FTIR spectroscopy (GAM-FTIR): applications to adhesion studies, *Surface and Interface Analysis*, **1999**, *27*, 865-870.
2. Umemura, J., Kamata, T., Kawai, T. and Takenaka, T. Quantitative evaluation of molecular orientation in thin Langmuir-Blodgett films by FT-IR transmission and reflection-absorption spectroscopy, *Journal of Physical Chemistry*, **1990**, *94*, 62-67.
3. Knoll, W., Philpott, M. R. and Golden, W. G. Surface infrared and surface enhanced Raman vibrational-spectra of monolayer assemblies in contact with rough metal-surfaces, *Journal of Chemical Physics*, **1982**, *77*, 219-225.
4. Kramer, R. *Chemometric techniques for quantitative analysis*; Marcel Dekker Inc: New York, 1998.

Chapter 6

SDS on a stainless steel substrate: the effect of surface roughness

6.1 Introduction

The research presented in the previous chapter clearly shows that IRRAS combined with PLS models can be used to identify and quantify SDS contamination on smooth aluminium surfaces. The physical and chemical properties of stainless steel mean that it is a superior material to aluminium in pharmaceutical and food manufacturing, and it is widely used for mixers, reactors, and work surfaces in these industries. So it is important to establish that identification and quantification of the type demonstrated in Chapter five can be verified for stainless steel. Furthermore, since the surfaces of manufacturing equipment are not necessarily uniformly smooth, it is also necessary to evaluate the effects of surface roughness on the quality of the predictions from the relevant chemometric models.

Theoretical treatments of surface roughness on IRRA spectra will not be covered in this thesis. Research has been carried out by Caron and Jacquet,[1] who used Kirchhoff theory[2] to account for variations in the IRRA spectra of organic films on roughened metallic substrates where the radiation wavelength is comparable to the dimensions of roughness, (detail on Kirchhoff theory, which is used to describe the scattering of both acoustic and electromagnetic waves scattering from roughened surfaces, can be found in a book by J. A. Ogilvy).[2] The effect of surface roughness on spectra collected from dielectric surfaces has also received some attention,[3] although those researchers only considered the roughening of the bottom surface of a plate where the sample film has been deposited on the substrate's flat top face.

Stainless steel is a more complicated material than aluminium because it is an alloy of iron, chromium and nickel with trace amounts of elements such as molybdenum and manganese. The exact composition depends on the grade of stainless steel. This means that different grades can have different refractive indices,[4] and therefore different IRRA spectra. The inference that the transfer of chemometric calibrations between grades may be difficult has not been investigated here, and only type 316 stainless steel was used in this research.

6.2 Sample preparation and characterisation

Seventy-two samples (23 on polished, 24 on smooth, and 25 on rough plates – see descriptions in Section 4.5) were prepared using the spray technique with solutions of SDS dissolved in Milli-Q water. Care was taken in selecting the sample plates to ensure that they were flat (no warps and bends) and free from surface distortions that could influence the baseline. By varying the concentration of the solution and the distance of the airbrush nozzle to the substrate surface, it was possible to achieve almost uniform surface loadings in the range 0–2.1 μg cm^{-2}. Higher loadings were not investigated since they are readily visible to the eye.

One single-channel background spectrum was collected from each of three clean substrate finishes. Eight single-channel sample spectra were collected from each sample surface, as described in Section 5.2. By combining sample spectra with background spectra collected from the same *and* different finishes, it was possible to investigate in more detail the way in which the finish influenced the calibration and whether it would be possible to build a combined model for all finishes.

The average surface loadings were determined from the aqueous washings using quantitative ^{1}H NMR (Section 4.9). The spectra from different surface finishes were similar, and a typical example (for a 2.07 μg cm^{-2} SDS on a polished stainless steel

substrate with polished stainless steel as the background) is shown in Figure 6.1. The SDS spectral regions of interest, indicated by the horizontal bars above the *x*-axis, 2798–3022 cm^{-1} and 1000–1302 cm^{-1}, are the same regions as used for SDS on aluminium.

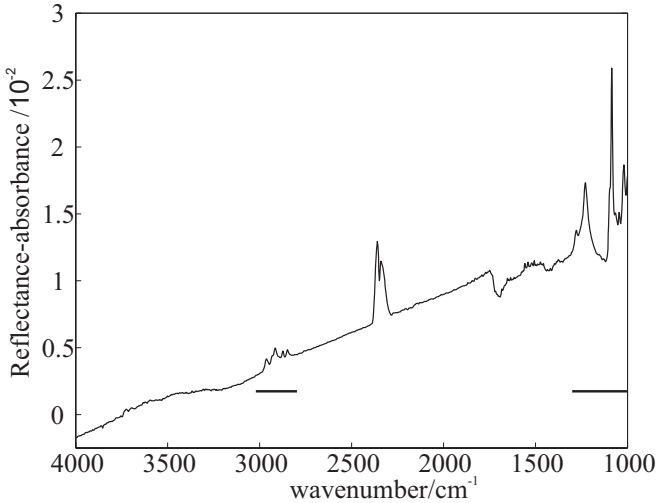


Figure 6.1: IRRA spectrum for an SDS surface loading of 2.07 μg cm^{-2} on polished stainless steel. The horizontal bars above the *x*-axis indicate the SDS bands that were used in the quantitative analysis.

SEM images of some samples are shown in Figure 6.2. These were taken in two orientations; the images on the left look vertically down on the surface, while those on the right are at ~80° to the normal, the same as the incidence angle of the IRRAS grazing-angle sampling head. The electron beam voltage was 10 kV for all images but the electron beam current (*I*) was varied depending on the surface finish and is shown next to each image. The schematic next to each image indicates the location of the SDS "spots".

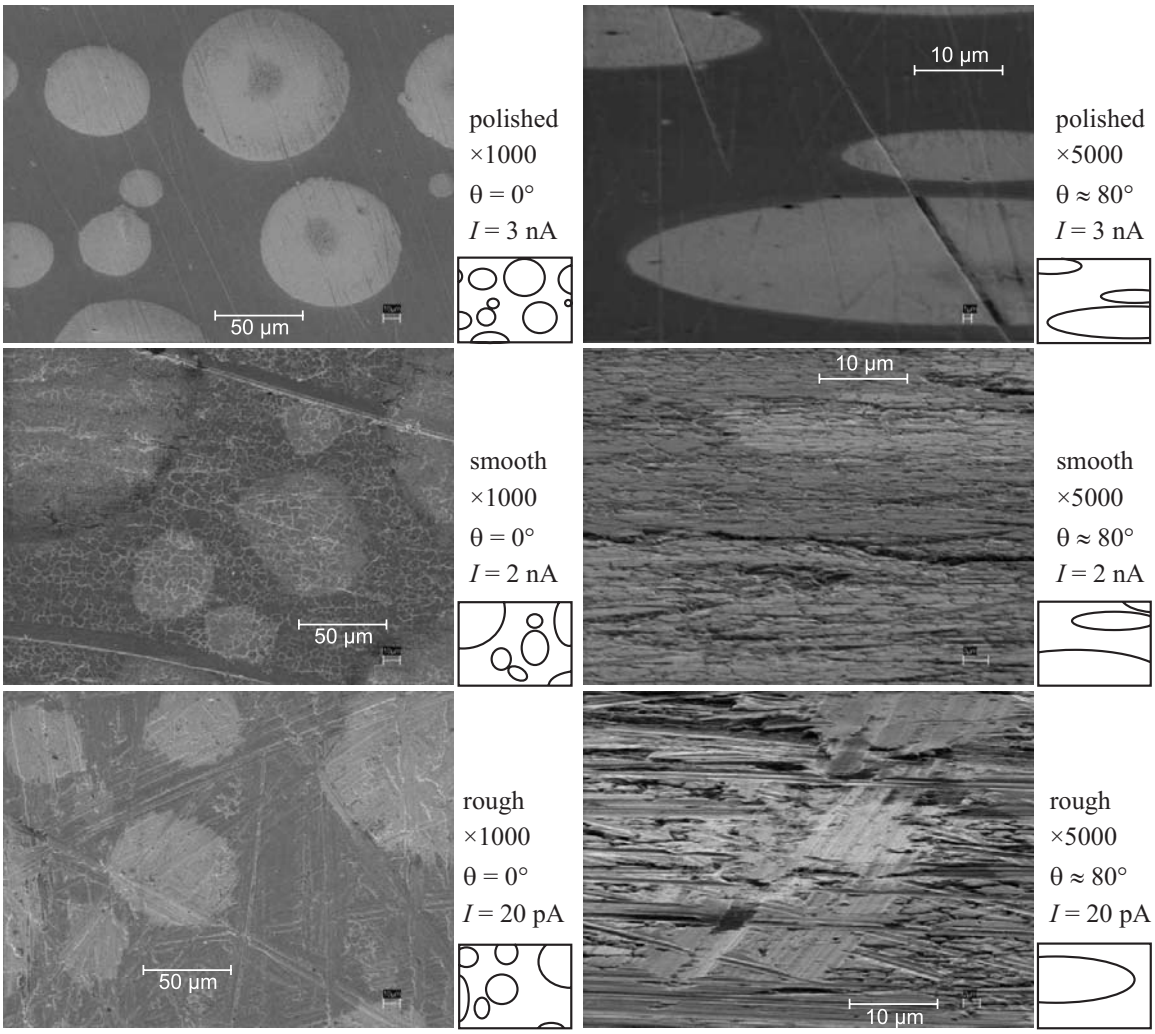


Figure 6.2: SEM images for SDS on polished, smooth and rough stainless steel substrates, the top row of images are from polished, second row are from smooth and the third row from rough. Images in the left column are taken looking vertically down on a flat surface at × 1000 magnification and in the right column are images taken looking across the surface at ~80° (the grazing-angle) at × 5000 magnification. The electron beam voltage is 10 kV for all images and the electron beam current is varied depending on the surface finish and is shown next to each image. The schematics next to each image details the location of the SDS "spots".

The SDS surface loading for the samples imaged by SEM were ~1 µg cm^{-2} for all surface finishes. This was determined using quantitative ^{1}H NMR after rinsing the 2 × 2 cm^2 plates with 3 mL of Milli-Q water. Care was taken to avoid contamination from the carbon pad used to stick the substrate to the stage of the SEM. Because of the small amount of material, the loadings so determined are not as reliable as those determined for the larger substrate samples.

From the SEM images and approximate surface loadings, the thicknesses of the SDS spots, *d*, was estimated using equation (6.1), where *c* is the loading, *A* is the substrate area, *a* is the area actually covered by the analyte (*a*/*A* is the fractional coverage) and ρ is the density of the analyte material. From Figure 6.2, $a/A \approx 0.4$. With $c \approx 1$ μg cm^{-2} and ρ = 1.1 g cm^{-3},[5] this gives $d \approx 20$ nm At this thickness, films are predicted (see Section 2.4.3) to display a quantitative linear relationship between IRRA intensity and surface loading.

$$d = \frac{c\,A}{\rho\,a} \tag{6.1}$$

The SEM images in Figure 6.2 reveal clear differences between the stainless steel finishes. The polished plates (top row of Figure 6.2) have areas that are flat (on the sub-micrometer scale) apart from relatively sparse, millimetre-length scratches that are separated laterally by tens of micrometers and have transverse dimensions (widths and depths) of ~1 μm. The smooth plates (middle row) have a much greater density (separations less than 1 μm) of microscopic scratches and folds, which are shorter (tens of micrometers) than the polished-surface scratches, but have similar lateral dimensions. Less common are larger and much deeper scores, which run at slight diagonals across the mid-row images of Figure 6.2. On the rough-surface plates, the deep, larger-scale scores are much more prevalent. These SEM images also show clear evidence that part of each droplet flows into the scratches and grooves prior to complete solvent evaporation; this gives rise to the jagged disk edges that are particularly apparent in the smooth- and rough-surface micrographs of Figure 6.2.

6.3 Quantitative analysis using PLS

To test whether a quantitative relationship exists between surface loading and IRRA spectra of SDS on stainless steel surfaces, individual PLS models were built for the spectra collected from each surface finish. Initially, the stainless steel surface used for the

background was the same as used for the sample substrate. In keeping with the conclusions for SDS on aluminium, no data pre-treatments were performed.

Models were built using the fingerprint (1000–1302 cm^{-1}) and CH-stretching (2798–3022 cm^{-1}) spectral ranges independently, as well as the combined ranges. Three averaging approaches were used, as described in Section 3.12: non-averaged calibration with averaged prediction, NA:A; averaged calibration with averaged prediction, A:A ;and non-averaged calibration with non-averaged prediction, NA:NA.

The optimum rank for each model was determined using the statistical and empirical methods described in Section 3.8, and was compared with the optimum rank selected through visual inspection. The statistical method is based on the *F*-test as detailed in Section 3.8. In the empirical method, the optimum rank is the one with the minimum $Crit_r$ value (Section 3.8) determined using equation (6.2), where r is the rank. Two ways of calculating s, denoted s_1 and s_2 (equations (6.3) and (6.4)) were investigated, but found to have no affect on the results. The ω factor can be adjusted as desired; higher values make Crit less sensitive to increasing RMSECV and favour lower rank models. Experiments by Martens and Dardenne suggested that $\omega = 0.05$ is reasonable in most cases.[6] However, during this research it was found that $\omega = 0.05$ often selected models with too few factors (i.e. too small rank), and that $\omega = 0.025$ gave more consistent results.

$$Crit_r = RMSECV_r + \omega rs \tag{6.2}$$

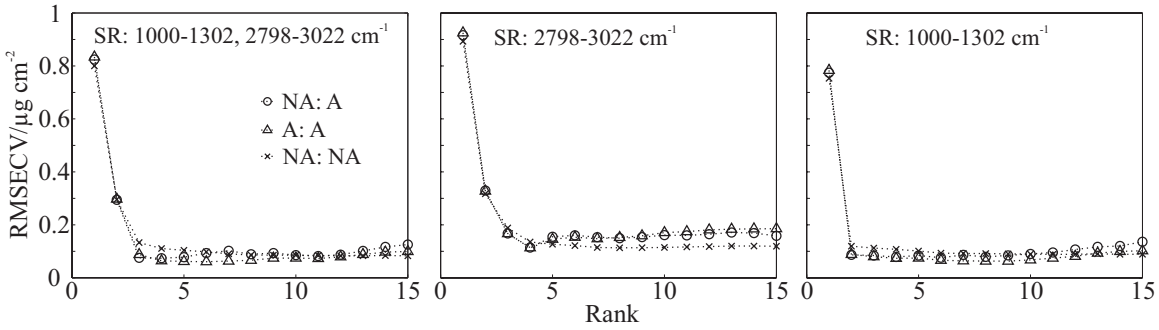
$$s_1 = RMSECV_0 \tag{6.3}$$

$$s_2 = \sqrt{RMSECV_0^2 + RMSECV_{r\,max}^2} \tag{6.4}$$

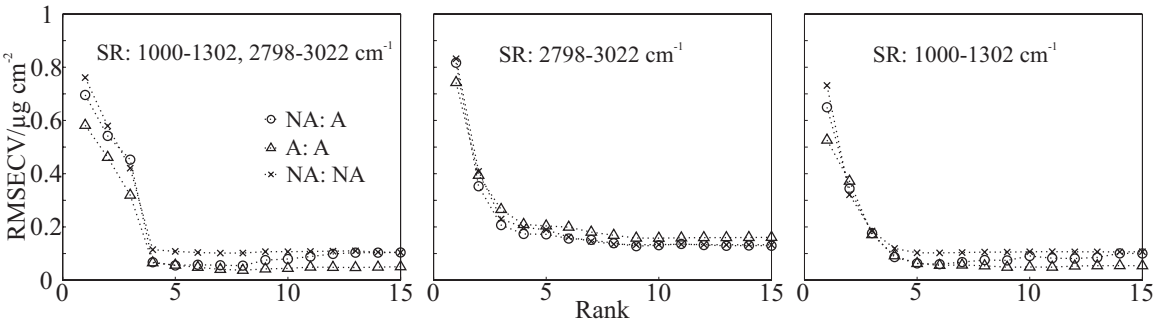
RMSECV *versus* rank plots for SDS on the three finishes of stainless steel are given in Figure 6.3 and the corresponding optimum ranks, according to the various determination methods, are presented in Table 6.1. The bottom two rows of each sub-table of Table 6.1 show the selected ranks and corresponding RMSECVs for each surface finish after all

methods had been considered. The predicted *versus* true loading plots for the NA:A models using the selected ranks are shown in Figure 6.4.

SDS on polished stainless steel:



SDS on smooth stainless steel:



SDS on rough stainless steel:

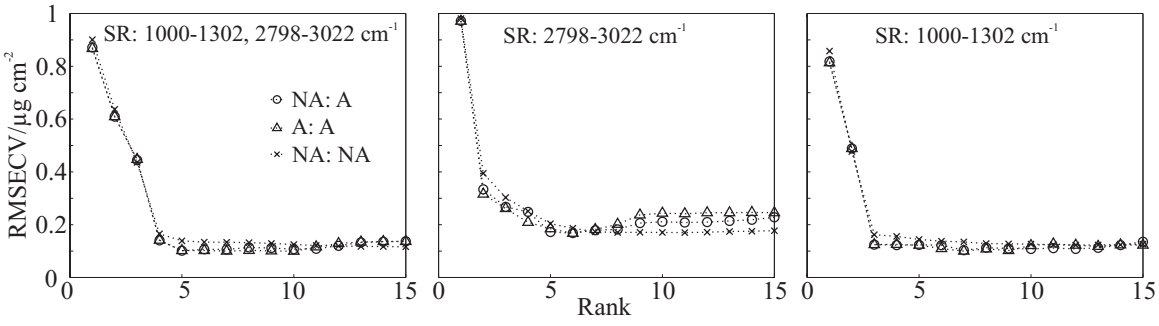


Figure 6.3: RMSECV *versus* rank plots for SDS on stainless steel. From top to bottom, the plots in each row pertain to polished, smooth and rough plates, using the same finish as the background. From right to left, the spectral ranges in each column are the combined range, CH-stretching range and fingerprint range. These are the same as were used in the SDS:aluminium analysis of Chapter five. The three data treatment approaches are: non-averaged calibration with averaged prediction, NA:A; averaged calibration with averaged prediction, A:A; and non-averaged calibration with non-averaged prediction, NA:NA.

Table 6.1: Selection of the optimum rank using *F*-test, $Crit_r$ and visual RMSECV methods for SDS samples prepared on polished (top), smooth (middle) and rough (bottom) stainless steel substrates.

SDS on polished stainless steel:

region:	combined			CH			fingerprint		
averaging approach:	NA:A	A:A	NA:NA	NA:A	A:A	NA:NA	NA:A	A:A	NA:NA
F-test α = 0.25	3	4	5	4	4	4	2	3	3
$Crit_r$ ω = 0.05	3	3	3	4	4	4	2	2	2
ω = 0.025	3	4	4	4	4	4	2	2	2
Visual REMSECV	3	3,4	4,5	4	4	4	2	2	2
selected rank	**3**	**4**	**4**	**4**	**4**	**4**	**2**	**2**	**2**
RMSECV at selected rank/ µg cm^{-2}	0.075	0.064	0.111	0.114	0.113	0.135	0.089	0.088	0.120

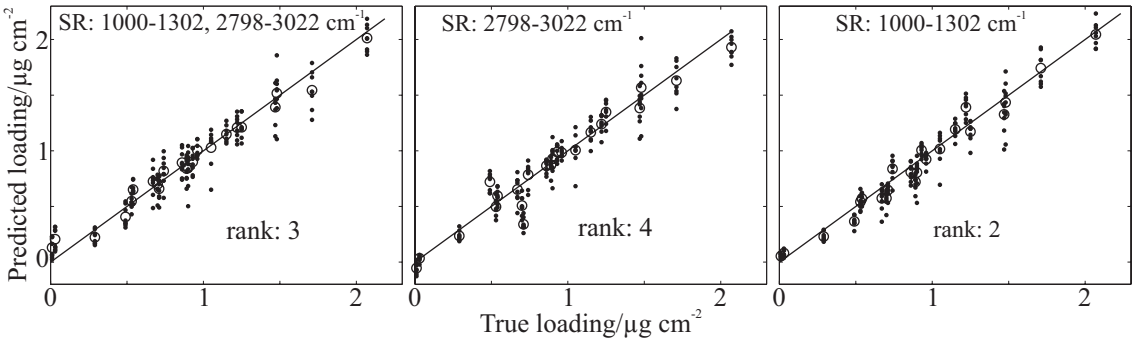
SDS on smooth stainless steel:

region:	combined			CH			fingerprint		
averaging approach:	NA:A	A:A	NA:NA	NA:A	A:A	NA:NA	NA:A	A:A	NA:NA
F-test α = 0.25	4	6	4	6	5	6	5	5	4
$Crit_r$ ω = 0.05	4	4	4	3	4	3	4	5	4
ω = 0.025	4	4	4	4	4	4	5	5	4
Visual REMSECV	3	4	4	4	4	4,6	5	5	4
selected rank	**4**	**4**	**4**	**4**	**4**	**4**	**5**	**5**	**4**
RMSECV at selected rank/ µg cm^{-2}	0.067	0.066	0.115	0.174	0.209	0.198	0.063	0.063	0.119

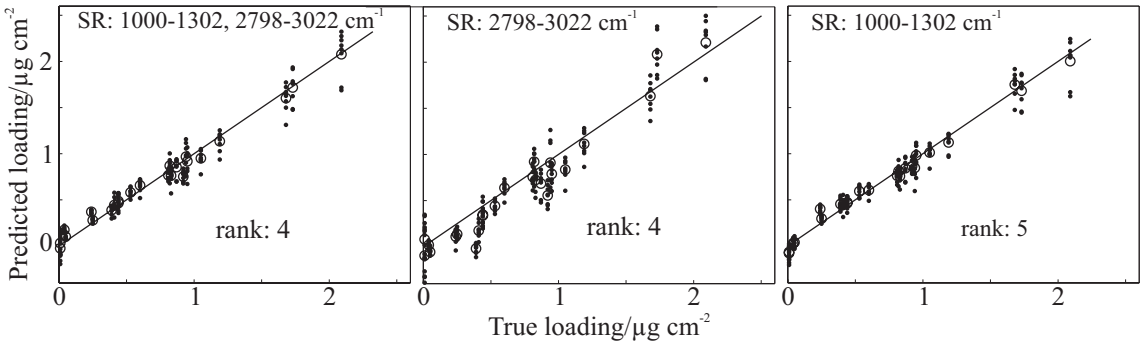
SDS on rough stainless steel:

region:	combined			CH			fingerprint		
averaging approach:	NA:A	A:A	NA:NA	NA:A	A:A	NA:NA	NA:A	A:A	NA:NA
F-test α = 0.25	5	5	5	5	4	5	3	3	4
$Crit_r$ ω = 0.05	4	4	4	3	4	4	3	3	3
ω = 0.025	5	5	5	5	4	5	3	3	3
Visual REMSECV	5	5	5	5	5,6	5,6	3	3	3
selected rank	**5**	**5**	**5**	**5**	**4**	**5**	**3**	**3**	**3**
RMSECV at selected rank/ µg cm^{-2}	0.101	0.102	0.139	0.172	0.209	0.204	0.124	0.125	0.162

SDS on polished stainless steel:



SDS on smooth stainless steel:



SDS on rough stainless steel:

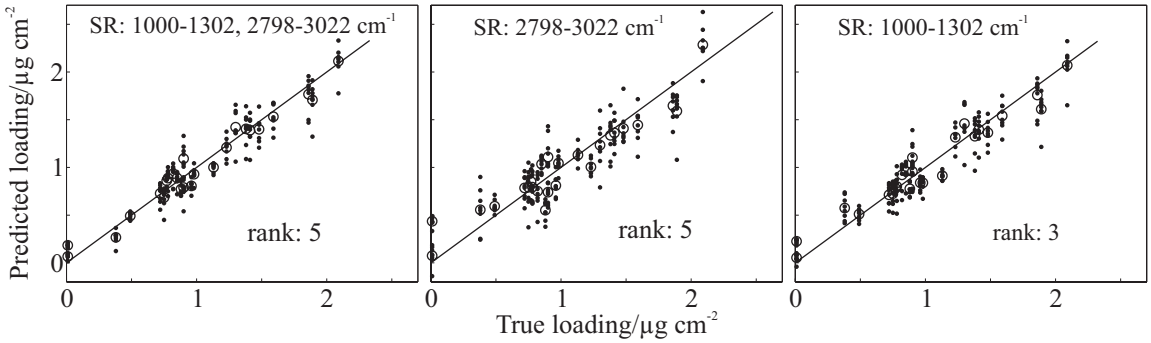


Figure 6.4: The predicted *versus* true loading plots for the NA:A models using the selected ranks in Table 6.1. The dots represent the values for the non-averaged spectra and the open circles the averaged results for a single sample. From top to bottom, the plots in each row pertain to polished, smooth and rough plates, using the same finish as the background. From right to left, the spectral ranges in each column are the combined range, CH-stretching range and fingerprint range.

6.3.1 Comparison between all substrate finishes

The selected optimum ranks and the corresponding RMSECV values for all roughness models have been gathered into Table 6.2.

Table 6.2: The selected optimum ranks from Tables 6.1 for all roughness models, along with their corresponding RMSECV values in μg cm^{-2}.

region	model data treatment	substrate:	polished	smooth	rough
combined	NA:A	rank	3	4	5
		RMSECV	0.075	0.067	0.101
	A:A	rank	4	4	5
		RMSECV	0.064	0.066	0.102
	NA:NA	rank	4	4	5
		RMSECV	0.111	0.115	0.139
CH	NA:A	rank	4	4	5
		RMSECV	0.114	0.174	0.172
	A:A	rank	4	4	4
		RMSECV	0.113	0.209	0.209
	NA:NA	rank	4	4	5
		RMSECV	0.135	0.198	0.204
fingerprint	NA:A	rank	2	5	3
		RMSECV	0.087	0.063	0.124
	A:A	rank	2	5	3
		RMSECV	0.088	0.063	0.125
	NA:NA	rank	2	4	3
		RMSECV	0.120	0.119	0.162

The conclusions drawn from Table 6.2 are that, in general, the best results (i.e. smallest ranks and RMSECVs) are seen in models built using the polished samples and the worst results are from the rough samples. This trend is likely to be due to the homogenous nature of the surface finish on the scale of the radiation; the polished surface is almost

devoid of surface features whereas the rough surface has many features over a range of dimensions (see Figure 6.2).

A comparison of the optimum rank selection methods shown in Table 6.1 reveals that the different selection methods have different sensitivities. The *F*-test can be very sensitive to increases in the RMSECV and generally selects optimum ranks greater than, or equal to, those selected by the other methods. The sensitivity could be reduced by adjusting α, but this was not carried out. In the $Crit_r$ method, the adjustment factor, ω, has a reasonable influence on the optimum rank. It was generally found that $\omega = 0.05$ makes the method insensitive to increasing RMSECV values, resulting in a calibration with too few factors, which underfits the data. More consistent optimum ranks, in line with those selected by the other methods were achieved with $\omega = 0.025$.

The calibrations prepared for SDS on polished and rough stainless steel finishes display similar trends to those seen for SDS on aluminium, with the CH-stretching region being less well behaved than the fingerprint region, and the combined ranges giving rise to a higher optimum rank than the fingerprint region but with smaller RMSECV values.

The overall trends seen for SDS on smooth stainless steel at the selected optimum ranks are different to those for the polished and rough plates. The main difference is that optimum ranks for the fingerprint region are greater than those for the CH-stretching and combined-region models. However, if the RMSECV values for the optimum ranks are considered, the results are more similar to those of the other finishes, since both the combined regions and fingerprint region models have similar RMSECV values whereas the RMSECV values for the CH models are substantially greater. In this situation it is easily concluded that the two spectral regions contain complementary information since the combined spectral region models combine small ranks with small RMSECV values.

6.3.2 Determination of outliers

A model with too few factors (i.e. an underestimated optimal rank) will underfit the calibration data, leading to a poor model that does not utilise all of the available information. On the other hand, too many factors causes overfitting where "noise" is incorporated into the calibration, also leading to a poor calibration. Outlier spectra in the calibration set can have large influences on the calibration model and the selected rank and should therefore be excluded from the calibration. To detect any outliers, the leverages and standardised loading residuals were determined for the optimum PLS models using the combined spectral regions and the NA:A averaging approach. As shown in Figure 6.5 there are several suspect spectra (whose data are enclosed by dashed circles) with residuals greater than three standard deviations. After investigation, none of the corresponding samples were excluded from the calibrations because the offending leverages were relatively small ($h < 0.04$) and the rest of the spectra from the same samples had small leverages and small residual values. The decision not to exclude any data is consistent with a suggestion that leverage scores exceeding $2n/m$ or $3n/m$ (where m is the number of samples and n is the number of wavelengths) constitute influential outliers.[7] Taking a conservative approach where m is the number of spectra (rather than the number of samples), $2n/m \approx 3$ for all of the finishes, which is much greater than any of the leverages in Figure 6.5.

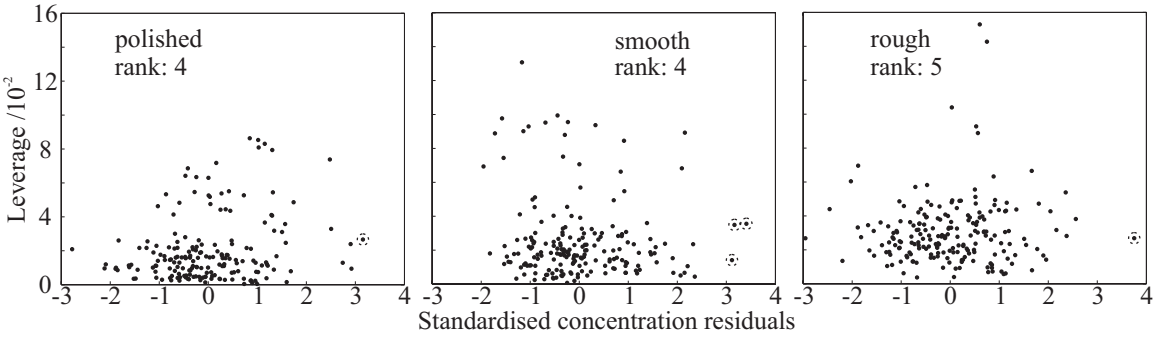


Figure 6.5: The leverage *versus* standardised residuals for all spectra in the calibration sets for the polished, smooth and rough samples.

6.4 The effect of background surface roughness on the PLS calibration

With the collection of background single-channel spectra from each of the finishes, the effect of altering the background finish can be investigated by recalculating the RA spectra using alternative background channels.

A major difference between the background single-channels lies with their intensities; the smoother the surface finish the greater the amount of radiation reflected to the detector. In comparison with the maximum of the ADC (analogue to digital converter) count of the spectrometer, the average signals from the three clean substrate surfaces were: polished, ~90%; smooth, ~85%; and rough ~76%. They also have wavelength dependent components, the effects of which can be seen in Figure 6.6, where RA spectra are shown for a sample single channel from a polished stainless steel surface (SDS loading of 1.15 μg cm^{-2}) combined with each of the three background single-channels. Similar effects are seen for SDS on either smooth or rough samples surfaces (Figure 6.7). Generally, and not surprisingly, when the background and sample finishes are the same, the RA baseline is closer to zero and has a lesser slope.

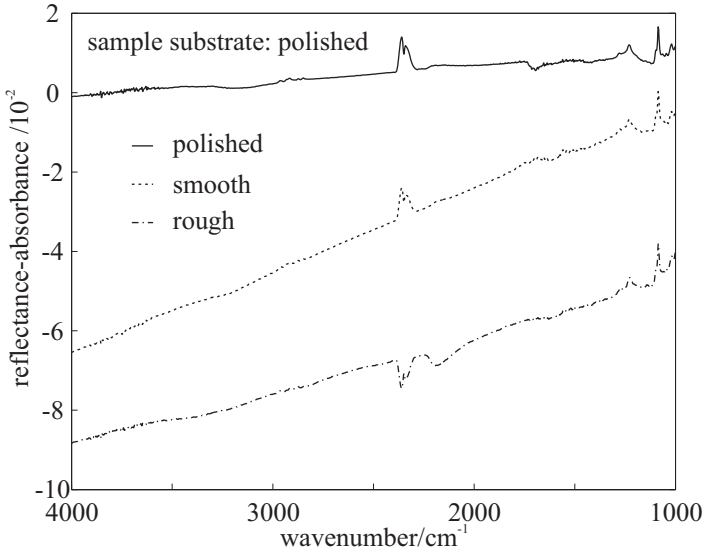


Figure 6.6: RA spectra for a sample single-channel with a SDS surface loading of 1.15 μg cm^{-2} on a polished stainless steel surface, calculated with polished, smooth and rough background single-channels.

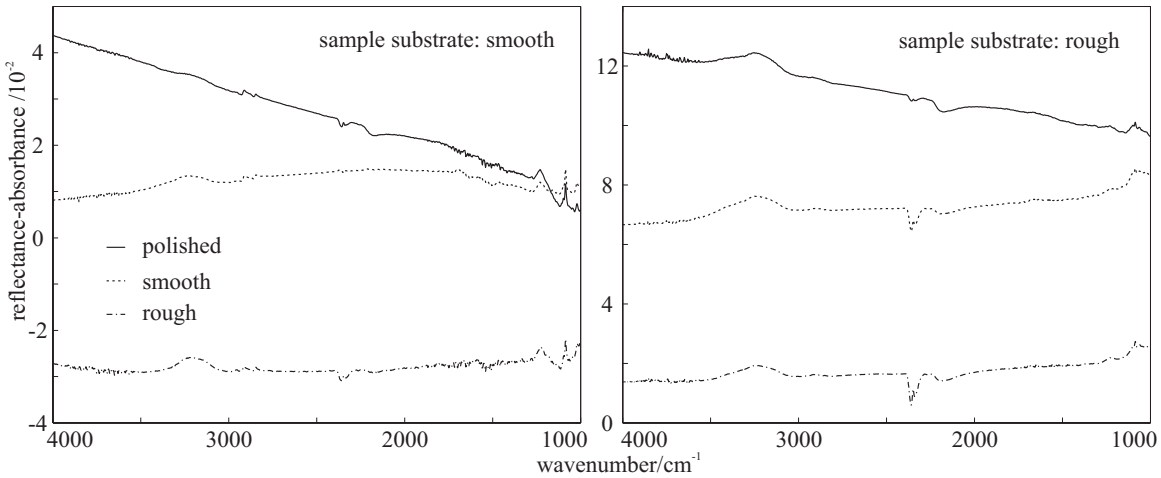


Figure 6.7: RA spectra of SDS on stainless steel for sample single-channels calculated with different background single channels Left: For a SDS surface loading of 1.73 μg cm^{-2} on smooth stainless steel; Right: For a SDS surface loading of 1.56 μg cm^{-2} on rough stainless steel.

These results suggest that PLS models built for calibration sets where the background and sample surface finishes are different may require extra factors (higher optimum ranks) to cope with offset and sloping baselines. The results obtained (using the previously outlined criteria) over the combined spectral range for all combinations of backgrounds are shown in Table 6.3. The entries in the shaded cells, which relate to cases where the sample and background finishes are the same, were previously presented in Tables 6.1 and 6.2.

Table 6.3: The selected optimum ranks for models built using different surface finishes for the background single-channels along with their corresponding RMSECV values in μg cm^{-2}.

spectral region: sample surface	combined averaging		background surface		
			polished	smooth	rough
polished	NA:A	rank	3	4	4
		RMSECV	0.075	0.083	0.086
	A:A	rank	4	4	4
		RMSECV	0.064	0.080	0.081
	NA:NA	rank	4	4	4
		RMSECV	0.111	0.105	0.109
smooth	NA:A	rank	5	4	5
		RMSECV	0.075	0.067	0.114
	A:A	rank	4	4	5
		RMSECV	0.080	0.066	0.111
	NA:NA	rank	5	4	5
		RMSECV	0.118	0.116	0.141
rough	NA:A	rank	5	6	5
		RMSECV	0.120	0.112	0.101
	A:A	rank	5	5	5
		RMSECV	0.119	0.121	0.102
	NA:NA	rank	5	5	5
		RMSECV	0.142	0.142	0.139

As anticipated, the best PLS calibrations are obtained when the sample and background finishes are the same. This is judged on the basis that such cases have the smallest optimum ranks and the lowest RMSECV values. Although the models calculated using the differing background finishes still gave surprisingly good results, in most cases the RMSECV values were $\lesssim 0.02$ μg cm^{-2} different from those for the same sample and background surface finish. The reason for these good results may be due to spectra in a

calibration having similarly sloping baselines, which may be compensated for by incorporation into a single PLS factor.

The conclusions from this section are that PLS calibrations can be prepared for SDS on stainless steel substrates even in situations where the background and sample surface finishes are different. However, better calibration models (smaller rank and RMSECV) are generally prepared when the same finish is used for both.

6.5 Combined models: two surface finishes

Following the success of the PLS calibrations for SDS films prepared on single substrate finishes, the possibility of a combined PLS calibration was investigated. This is the case where a PLS model could be built to predict the SDS surface loading on any surface finish using backgrounds from a single surface finish.

Two sample sets were generated; the first combined the samples from the polished and the smooth surfaces, while the second combined the smooth and the rough samples. RA spectra could then be calculated for these calibration sets using polished, smooth or rough background substrates. PLS models were calculated using the combined (2798–3022 cm^{-1} plus 1000–1302 cm^{-1}) spectral regions. The optimum ranks were selected using the previously described techniques and are indicated with solid symbols on the appropriate RMSECV *versus* rank plots.

6.5.1 Combined polished and smooth model

The RMSECV *versus* rank plots for samples combined from polished and smooth surfaces using either the polished or the smooth background surfaces are shown in Figure 6.8. The shape of these plots are different from those for the individual calibrations; instead of flattening out abruptly, they follow a shallow curve that continues to decrease, making the optimum rank selection more difficult and subjective. This is especially

apparent when smooth substrates are used for the background, suggesting that better results are likely to be achieved when polished backgrounds are used. The selected optimum ranks are indicated by the filled symbols in Figure 6.8 and are presented with the corresponding RMSECV values in Table 6.4.

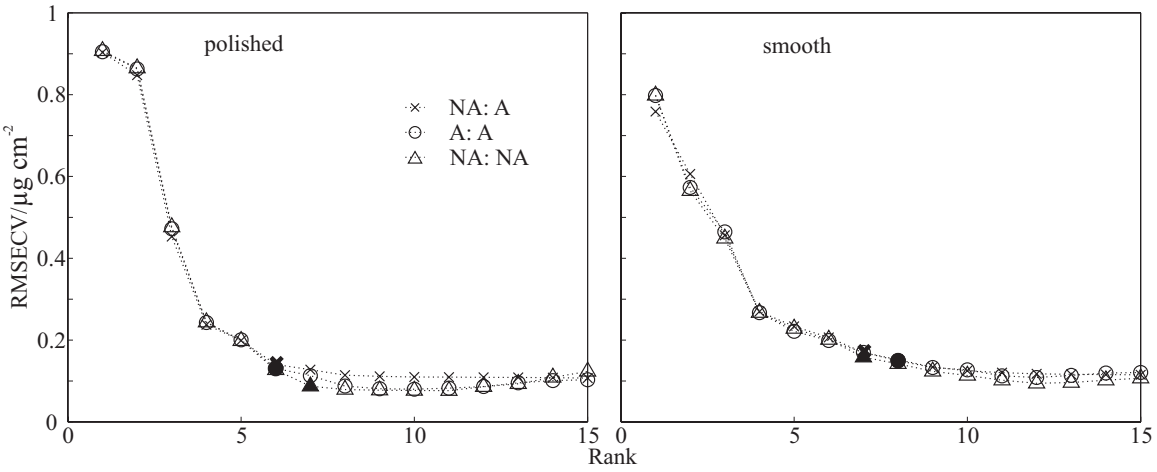


Figure 6.8: RMSECV *versus* rank plots for combined data from SDS on polished and smooth stainless steel plates using either polished (left) or smooth (right) clean stainless steel plates as the background. The combined SDS spectral regions were used in the PLS analysis. The three data treatment approaches are: non-averaged calibration with averaged prediction, NA:A; averaged calibration with averaged prediction, A:A; and non-averaged calibration with non-averaged prediction, NA:NA. The optimum rank for each data treatment is indicated by a solid symbol.

Table 6.4: The selected optimum ranks and corresponding RMSECV values in µg cm^{-2} for PLS calibrations built using the combined calibration set with SDS films prepared on polished and smooth surface finishes.

spectral regions:	2798–3022 cm^{-1} plus 1000–1302 cm^{-1}		background surface	
sample surfaces	averaging approach		polished	smooth
polished and smooth	NA:A	rank	6	8
		RMSECV	0.131	0.150
	A:A	rank	7	7
		RMSECV	0.087	0.158
	NA:NA	rank	6	7
		RMSECV	0.140	0.170

The predicted *versus* true loading plots at the optimum ranks for the NA:A averaging approach are shown in Figure 6.9. Only the average prediction for each sample is plotted to reduce congestion within the plots. The samples from the two stainless steel finishes are shown with different symbols. From a visual inspection, it appears that there is no obvious bias in the determination of either the polished or smooth sample loadings. This is a positive result, but the overall results are not as good as those for the individual substrate finishes since the optimum ranks are larger, as are the associated RMSECV values.

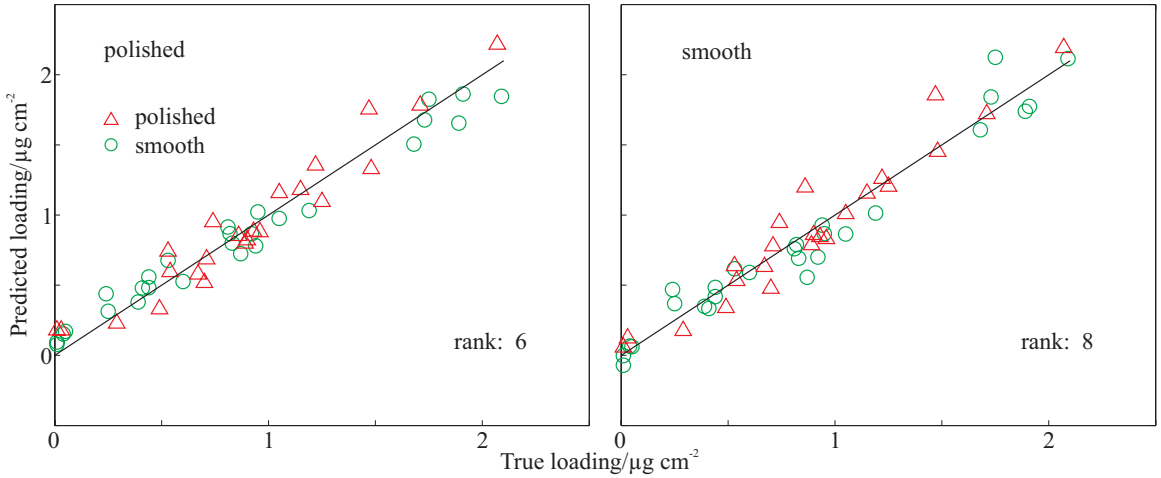


Figure 6.9: Predicted *versus* true loading plots for the NA:A models using the optimum ranks selected in Figure 6.8. The plots correspond to the different substrate finishes used for the background single-channel (left: polished, right: smooth). To reduce congestion, only the average prediction for each sample is shown. The data for polished and smooth substrate finishes are indicated by triangles and circles, respectively.

6.5.2 Combined smooth and rough model

The second of the investigated combined systems involved smooth and rough surfaces. On a macroscopic level, the surface finishes of the polished and smooth surfaces look and feel more similar to each other than either does to the rough finish. But the SEM images (Figures 6.2 and 6.10) change this opinion. The smooth surface has a “pancake” like appearance, probably caused by the hot rolling process used to form the stainless steel sheets, and the SEM images of the rough and smooth surfaces appear more similar than the polished. It was therefore suspected that better results might be achieved by the combined smooth and rough samples than the combined polished and smooth samples.

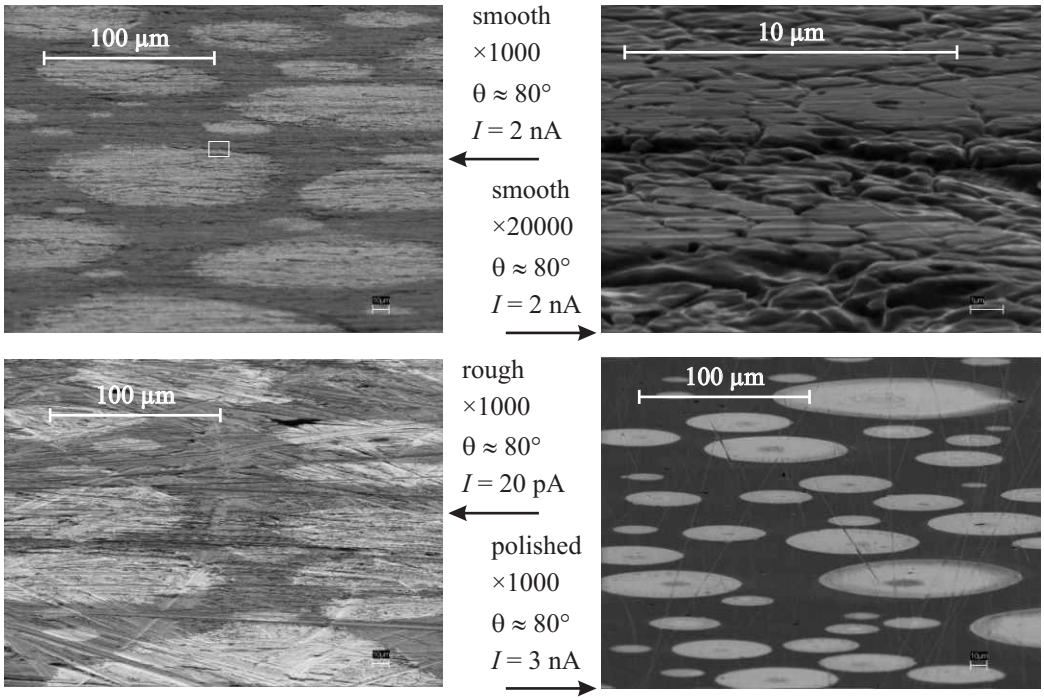


Figure 6.10: SEM images for SDS films on polished, smooth and rough stainless steel substrates. The top images are from smooth surfaces, with the white box in the centre of the left image enclosing the area that is enlarged in the image to the right. The bottom images are from rough (left) and polished (right) surfaces. All were taken at ~80° to the surface normal. The electron beam voltage was 10 kV and the electron beam current was varied depending on the surface, as shown next to each image.

The RMSECV plots for the combined smooth and rough sample data with smooth and rough background surfaces are shown in Figure 6.11. The shape of these plots more closely resembles those found for the individual calibrations than was the case for the polished and smooth data. The RMSECV flattens out rather abruptly at the optimum rank. The optimum ranks and their corresponding RMSECV values are presented in Table 6.5.

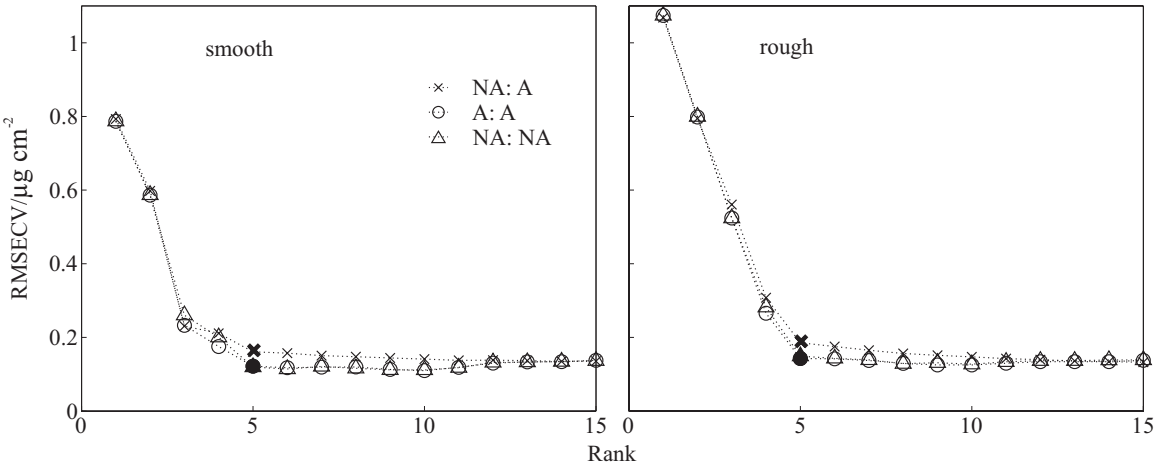


Figure 6.11: RMSECV *versus* rank plots for SDS films prepared on smooth and rough stainless steel plates using either smooth (left) or rough (right) clean stainless steel plates as the background. The combined SDS spectral regions were used in the PLS analysis. The three data treatment approaches are: non-averaged calibration with averaged prediction, NA:A; averaged calibration with averaged prediction, A:A; and non-averaged calibration with non-averaged prediction, NA:NA. The optimum rank for each data treatment is indicated by a solid symbol.

Table 6.5: The selected optimum ranks from Figure 6.11 and their corresponding RMSECV values in µg cm^{-2} for PLS calibrations built using the combined calibration set with SDS films prepared on smooth and rough surface finishes.

spectral regions:	2798–3022 cm^{-1} plus 1000–1302 cm^{-1}		background surface	
sample surfaces	averaging approach		smooth	rough
smooth and rough	NA:A	rank	5	5
		RMSECV	0.122	0.143
	A:A	rank	5	5
		RMSECV	0.120	0.149
	NA:NA	rank	5	5
		RMSECV	0.161	0.185

The results are similar to those for the individual calibrations; the optimum ranks are comparable and the corresponding RMSECV values are only slightly higher.

Visual inspection of the predicted *versus* true plots for the NA:A data treatment (Figure 6.12) indicates no obvious bias in the predictions for the smooth or rough samples.

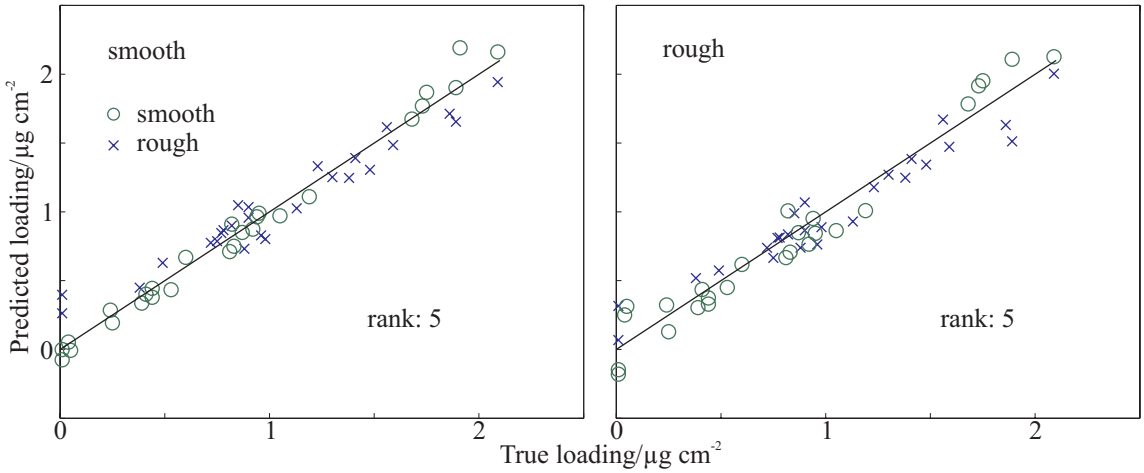


Figure 6.12: Predicted *versus* true loading plots for the NA:A models using the optimum ranks selected in Figure 6.11. The plots correspond to the different substrate finishes used for the background single-channel (left: smooth, right: rough). To reduce congestion, only the average predictions for each sample are shown. The samples prepared on smooth and rough substrate finishes are indicated by circles and crosses, respectively.

6.6 Combined model: three surface finishes

After the success of the PLS calibrations for SDS on single substrate finishes and combined models with two finishes, the possibility of a combined PLS calibration for all surface finishes was investigated. The initial study used the combined spectral regions (2798–3022 cm^{-1} plus 1000–1302 cm^{-1}), and the resulting RMSECV *versus* rank plots are shown in Figure 6.13 for the different backgrounds. The optimum ranks were selected using the previously described techniques and are indicated by solid symbols. This was reasonably straightforward for the polished and rough backgrounds, but when using the smooth backgrounds the optimum rank was very sensitive to the arbitrary $Crit_r$ factor.

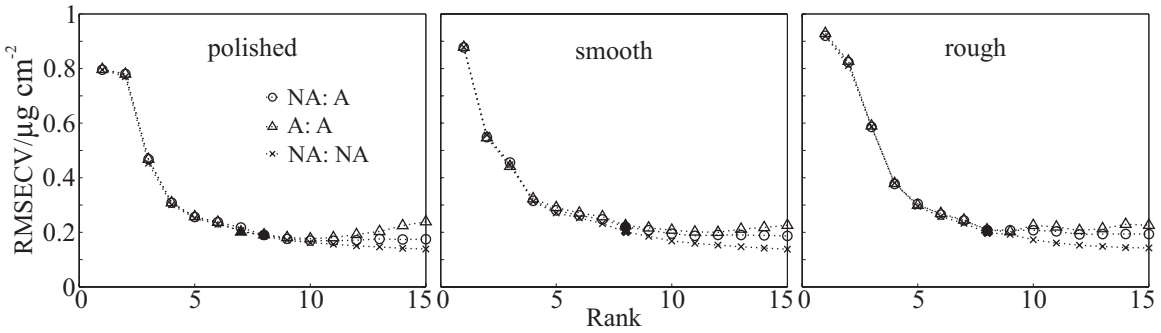


Figure 6.13: RMSECV *versus* rank plots for the combined calibration set of SDS on all stainless steel finishes. The plots correspond to each of the three clean stainless steel substrates used for the background single-channel, as shown at the top of each plot. The PLS models were generated using the combined spectral regions. The three data treatment approaches are: non-averaged calibration with averaged prediction, NA:A; averaged calibration with averaged prediction, A:A; and non-averaged calibration with non-averaged prediction, NA:NA. The optimum rank for each data treatment is indicated by a solid symbol.

As expected, the models using samples prepared on all surface finishes are poorer than those for individual calibrations; Figure 6.13 and Table 6.6 show that the optimum ranks are higher (~8 compared with ~4) as are the RMSECV values (~0.2 compared with ~0.1). The shapes of the RMSECV plots do not flatten out as abruptly, making the optimum rank selection more difficult and subjective. The overall results indicate that a PLS model can be generated, but at the cost of larger ranks and RMSECV values. The RMSECV values for each background are similar, although (like the two-finish combined models) better models are achieved with the polished background. The lowest RMSECV values are found for the NA:NA averaging approach, which is not unexpected since only one spectrum is removed during cross validation, leaving seven spectra from the same sample in the calibration. For this reason, the NA:NA results should be regarded with scepticism and have been included only for completeness.

Table 6.6: The selected optimum ranks from Figure 6.13 and their corresponding RMSECV values in μg cm^{-2} for PLS calibrations built using the combined calibration set with SDS film prepared on different surface finishes.

spectral regions:	2798–3022 cm^{-1} plus 1000–1302 cm^{-1}		background surface		
sample surface	averaging approach		polished	smooth	rough
combined	NA:A	rank	8	8	8
		RMSECV	0.191	0.220	0.208
	A:A	rank	7	8	8
		RMSECV	0.201	0.225	0.209
	NA:NA	rank	8	8	8
		RMSECV	0.185	0.202	0.199

A model was also generated for the combined data in which the individual spectra were calculated using matched background and sample finishes. The results, shown in Figure 6.14, are similar to those achieved when using the same background substrate for all samples. Since there appears to be no statistically compelling advantage to this type of analysis, and because it would be difficult to implement in a practical industrial environment (it would require several clean background substrates, to be matched with the sample finishes prior to sampling), this was not pursued further.

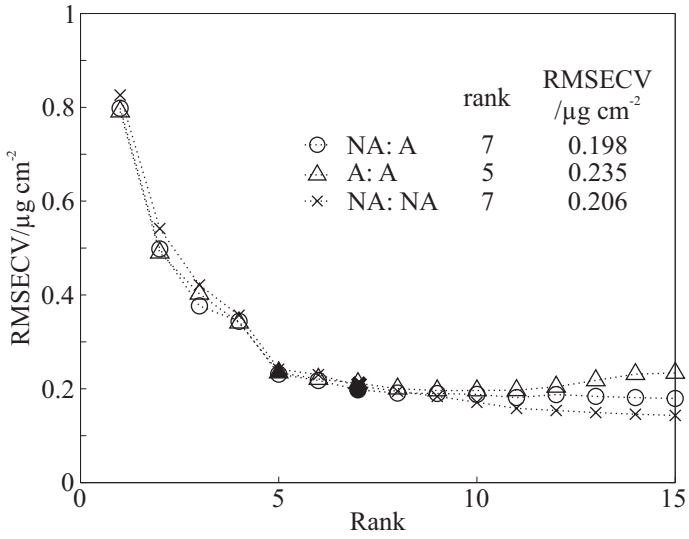


Figure 6.14: RMSECV *versus* rank plots for the combined calibration set with SDS on all stainless steel finishes. The background and sample single-channels are the same for each spectrum. The PLS models were generated using the combined spectral regions. The three data treatment approaches are: non-averaged calibration with averaged prediction, NA:A; averaged calibration with averaged prediction, A:A; and non-averaged calibration with non-averaged prediction, NA:NA. The optimum rank for each data treatment is indicated by a solid symbol.

The predicted *versus* true plots for the optimum ranks selected using the NA:A data averaging approach are shown in Figure 6.15; to reduce congestion and emphasise trends, only the averaged prediction for each sample is shown.

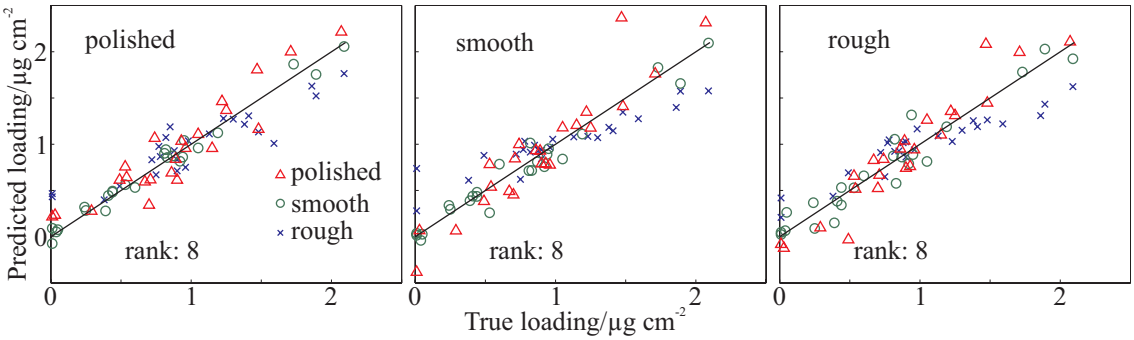


Figure 6.15: The predicted *versus* true loading plots for the NA:A models using the optimum ranks indicated in Figure 6.13. The three plots correspond to the substrate finishes used for the background single-channel. Only the average prediction for each sample is shown to reduce congestion within the plots. The samples prepared on polished, smooth and rough substrate finishes are indicated with triangles, circles and crosses respectively.

Although the plots in Figure 6.13 suggest that a PLS calibration can be prepared, the predicted *versus* true loading plots in Figure 6.15 indicate a more complicated situation

with systematic biases that depend on the surface finish. The predicted loadings for samples on polished substrates (indicated by triangles) are overestimated at higher loadings and underestimated at lower loadings, whereas the situation for the rough finishes (crosses) is the opposite; higher loadings are underestimated while lower loadings are overestimated. These trends are even more evident in Figure 6.16, which relates to a combined calibration set using a smooth background substrate but where the rank has been reduced to five, the optimum value selected with a larger $Crit_r$ factor. In addition, the predicted loadings for the smooth samples (circles) are all underestimated.

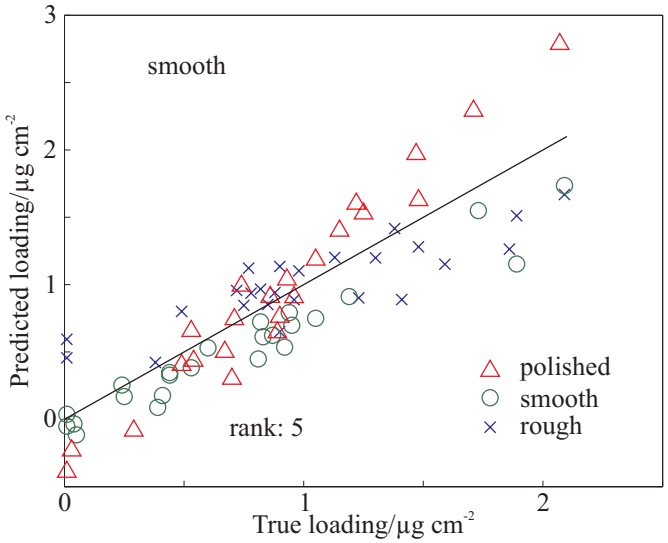


Figure 6.16: Predicted *versus* true loading plots for the NA:A model using smooth substrate finishes as the background. The model rank is smaller than the optimum rank selected earlier and was selected using a less sensitive $Crit_r$ factor.

These results indicate that, even though PLS can be used to build a calibration model to predict SDS surface loadings on stainless steel substrates independent of sample surface roughness, models extended over the whole range of finishes investigated here are unlikely to be robust and could lead to unreliable predictions. A possible reason for this problem could be the inclusion of both spectral regions. To investigate this further, PLS models were prepared using the CH and fingerprint regions independently, as described below.

6.6.1 CH spectral region

In Section 6.3 the individual surface roughness PLS models determined using just the CH spectral region were generally shown to be less well behaved than the models for either the combined or fingerprint regions. It was expected that this behaviour would influence the combined models. The RMSECV plots for the combined calibration set with the differing background surfaces, but using just the CH stretching (2798–3022 cm^{-1}) region, are shown in Figure 6.17, where the optimum rank for each averaging approach is indicated by a solid symbol.

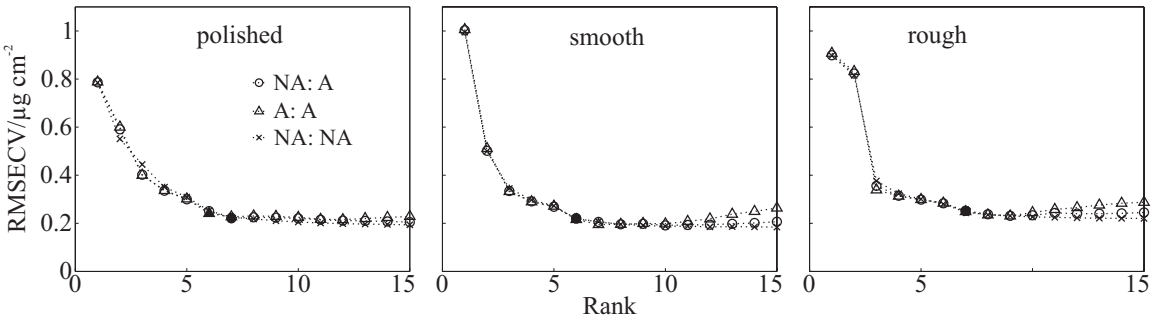


Figure 6.17: RMSECV *versus* rank plots for the combined calibration set of SDS on all stainless steel finishes. The three plots correspond to the stainless steel substrates used for the background single-channel shown at the top of each plot. The PLS models were determined using the CH spectral region. The three data treatment approaches are: non-averaged calibration with averaged prediction, NA:A; averaged calibration with averaged prediction, A:A; and non-averaged calibration with non-averaged prediction, NA:NA. The optimum rank for each data treatment is indicated by a solid symbol.

The curved shape of the RMSECV plots is consistent with combined spectral region models, although the selected optimum ranks are smaller. The optimum ranks and their corresponding RMSECV values are presented in Table 6.7. The predicted *versus* true loading plots for the NA:A averaging approach are shown in Figure 6.18. The trends in these plots are similar to those in the combined spectral region plots, but less pronounced; the most obvious similarity is that the rough sample loading predictions (crosses) are consistently underestimated at higher surface loadings.

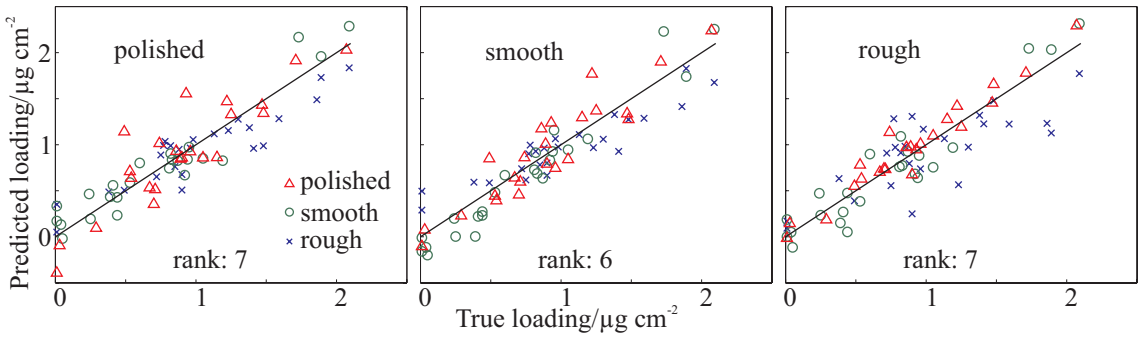


Figure 6.18: Predicted *versus* true loading plots for the NA:A models using the optimum rank selected in Figure 6.17. The three plots correspond to the substrate finishes used for the background single channel. Only the average prediction for each sample is shown to reduce congestion within the plots. The data for polished, smooth and rough surfaces are indicated with triangles, circles and crosses, respectively.

Table 6.7: The selected optimum ranks from Figure 6.17 and their corresponding RMSECV values in µg cm^{-2} for PLS calibrations built using the CH spectral region and the combined calibration set with SDS film prepared on different surface finishes.

spectral regions:	2798–3022 cm^{-1}		background surface		
sample surface	averaging approach		polished	smooth	rough
combined	NA:A	rank	7	6	7
		RMSECV	0.221	0.22	0.252
	A:A	rank	6	6	7
		RMSECV	0.241	0.216	0.249
	NA:NA	rank	7	6	7
		RMSECV	0.226	219	0.247

6.6.2 Fingerprint spectral region

Results presented in previous sections have shown that, in general, the best PLS models are achieved using the fingerprint region. The RMSECV plots for the combined sample calibration set with the differing background surfaces, but restricted to just the fingerprint region (1000–1302 cm^{-1}) of the spectrum, are shown in Figure 6.19, where the optimum rank for each averaging approach is indicated by a solid symbol.

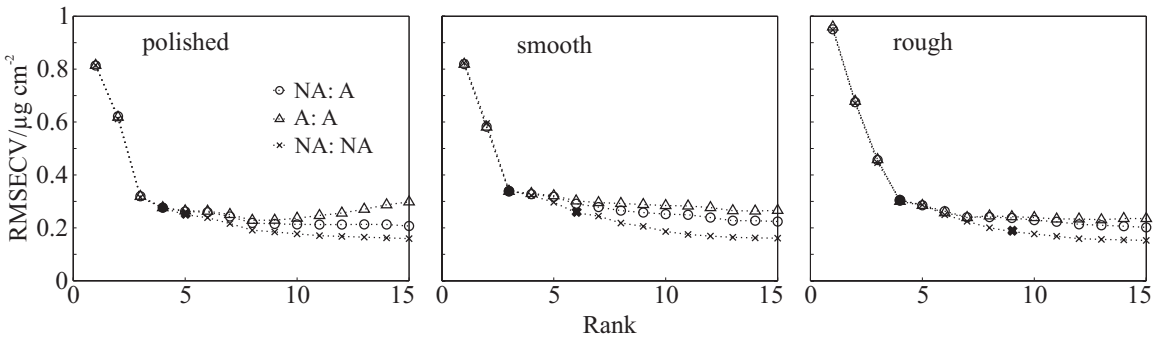


Figure 6.19: RMSECV *versus* rank plots for the combined calibration set with SDS films prepared on all stainless steel finishes. The three plots correspond to the clean stainless steel substrates used for the background single-channel, as shown at the top of each plot. The PLS models were calculated using the fingerprint (1000–1302 cm^{-1}) spectral region. The three data treatment approaches are: non-averaged calibration with averaged prediction, NA:A; averaged calibration with averaged prediction, A:A; and non-averaged calibration with non-averaged prediction, NA:NA. The optimum rank for each data treatment is indicated by a solid symbol.

The shapes of the RMSECV plots are much more in keeping with expectations described in Section 6.5.1 for a "good" model. The optimum ranks and associated RMSECV values are presented in Table 6.8. The optimum ranks are smaller than those for both the combined and CH spectral region models, although (unexpectedly) the associated RMSECV values are significantly higher. The predicted *versus* true loading plots for the NA:A averaging approach are shown in Figure 6.19 and the trends are of a type similar to, but stronger than, those present in the models for the combined spectral regions; the polished sample loadings (triangles) are overestimated at higher values and underestimated at lower values, most obviously when smooth or rough background substrates are used. The opposite trend is displayed by the rough sample predictions (crosses) and, to a lesser extent, by the smooth sample predictions (circles).

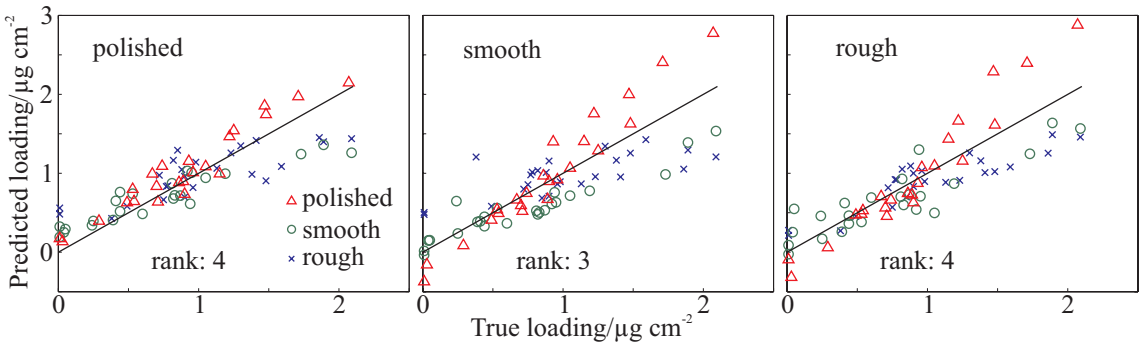


Figure 6.20: Predicted *versus* true loading plots for the NA:A models using the optimum rank shown in Figure 6.19. The three plots correspond to the substrate finishes used for the background single channel. Only the average prediction for each sample is shown this is to reduce congestion within the plots. The samples prepared on polished, smooth and rough substrate finishes are indicated by triangles, circles and crosses, repetitively.

Table 6.8: The selected optimum ranks from Figure 6.19 and their corresponding RMSECV values in µg cm^{-2} for PLS calibrations built using the fingerprint region and the combined calibration set with SDS film prepared on different surface finishes.

spectral regions:	1000–1302 cm^{-1}		background surface		
sample surface	averaging approach		polished	smooth	rough
combined	NA:A	rank	4	3	4
		RMSECV	0.276	0.339	0.304
	A:A	rank	4	3	4
		RMSECV	0.276	0.338	0.3
	NA:NA	rank	5	6	8
		RMSECV	0.25	0.257	0.2

6.6.3 Conclusions from the spectral region analysis

The overall conclusion from the analysis of PLS models using the individual spectral regions is that models built using a combined data set of SDS on polished, smooth and rough substrates do not lead to robust calibrations. The behaviour of the combined models suggests that there are differences between the spectra of samples prepared on different substrate finishes. This result is, perhaps, a little surprising considering the

results in Section 6.4 for the two-finish combined models. Why can apparently successful PLS calibrations be prepared for calibration sets for two substrate finishes, while calibration sets for three surface finishes lead to poor results? To answer this question, the spectra were investigated directly, using visual and univariate techniques.

6.7 Why do the combined sample calibrations break down?

The first step towards answering this question was to compare spectra collected from the three different substrate finishes with similar surface loadings. Three spectra were selected from the combined calibration set, one from polished, smooth and rough samples surfaces with loadings of 2.07, 2.09 and 2.09 μg cm^{-2}, respectively. Surface loadings at the higher end of the calibration range were selected since they have the most intense spectra, which should make it easier to detect changes in peak intensity and shape. Since the individual calibrations were successful it was expected that the spectra at higher loading would be representative of all those within the calibration set – this was confirmed by inspection of several other spectra. The spectra in the two regions used in the PLS calibrations are plotted in Figure 6.21.

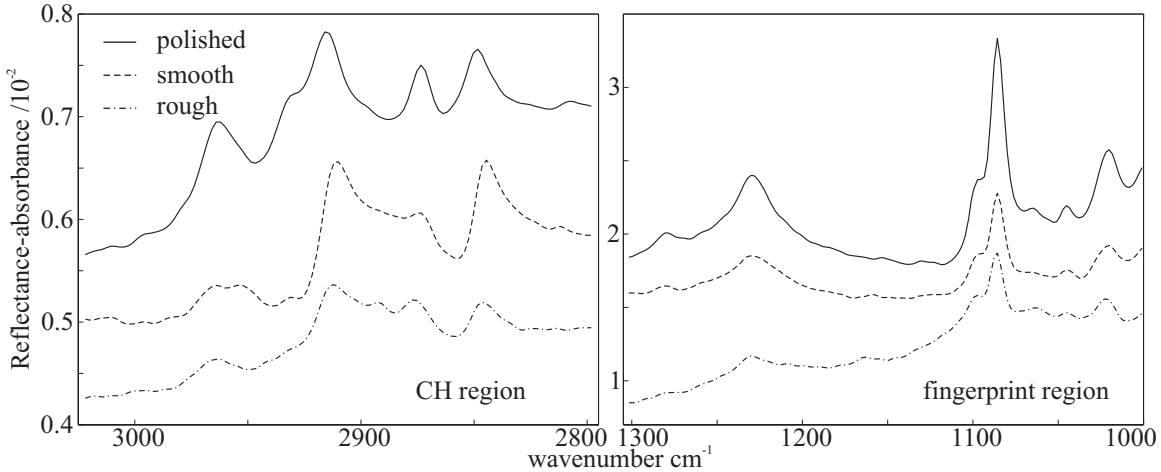


Figure 6.21: The spectra from the different substrate finishes with an SDS loading of ~2.1 μg cm^{-2}. The background single-channel was, in each case, taken from the same surface finish as the sample single-channel.

The most immediately obvious point in Figure 6.21 is that, despite similar loadings, the spectral intensities are quite different, especially in the fingerprint region. The spectrum from a polished substrate is much more intense than that from either of the smooth or rough substrates. There also appears to be a small intensity difference between the smooth and the rough spectra. In general, such differences are less evident in the CH region.

Even though the spectral intensities are different, the shapes of the spectra in the fingerprint region are very similar, but this is not true for the CH region. The most noticeable differences are the more intense bands at ~2870 and ~2930 cm^{-1} in the polished sample spectrum, the latter of which appears as a prominent shoulder on the blue (higher wavenumber) edge of the central peak. Other interesting differences are position shifts of the SDS features centred near ~2915 cm^{-1} and ~2850 cm^{-1}, which, in the polished spectrum, are shifted ~5 cm^{-1} to the higher wavenumbers (to the blue) of those in the smooth and rough spectra. Again, there are less prominent differences between the smooth and rough spectra; there is an additional band at ~2955 cm^{-1} in the smooth spectrum and the rough spectrum displays a band at ~2890 cm^{-1} (which may be present, but not as easily distinguished, in the polished and smooth spectra).

To be more quantitative about the intensity differences, the band areas were determined, as described in Section 5.3. Figure 6.22 shows these areas *versus* known surface loadings for the averaged spectra for each sample, with the bands for which the areas were determined being shown in the top row of plots. Since the shapes of the spectra in the CH region change with the finish, the area calculations were carried out over the entire range used in the PLS calibration using the integration method described in Section 5.3. In the case of the fingerprint region, the band positions are constant, which allowed the area calculations to be carried out over narrower spectral ranges, shown in Figure 6.22 and made the analysis less susceptible to influence from baseline effects. Clearly, the most intense spectral features are from polished surfaces and there are smaller differences between the smooth and rough intensities.

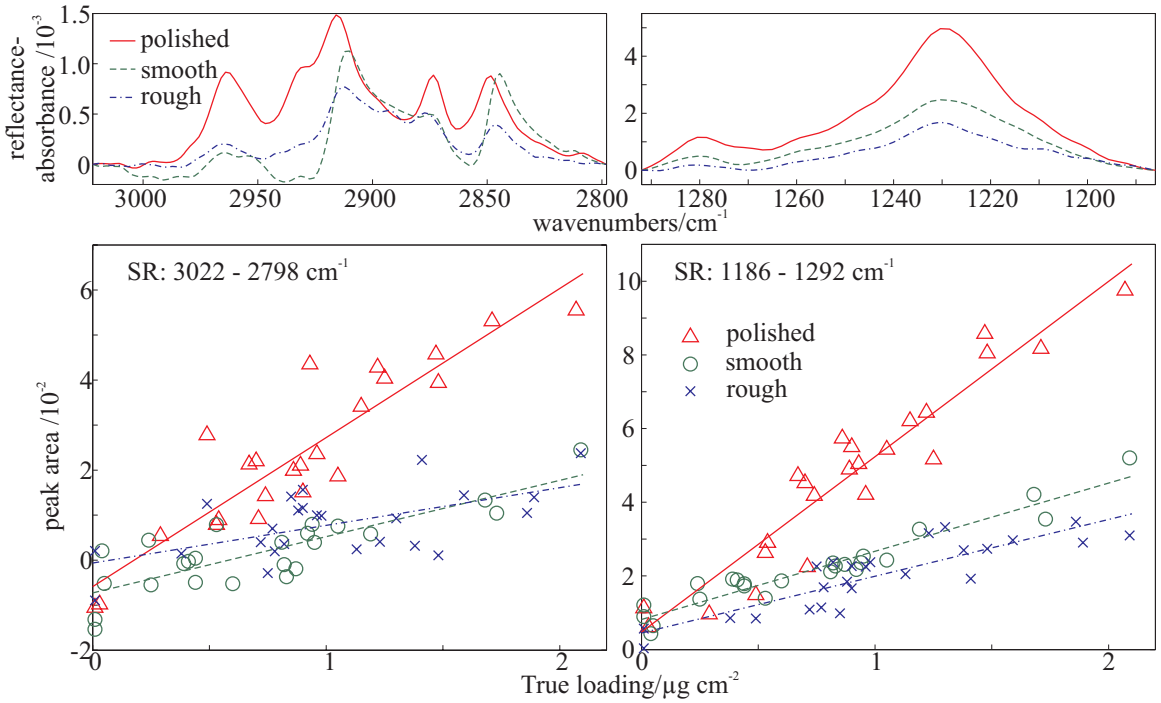


Figure 6.22: The lower plots are the band integrals *versus* surface loadings for the spectra shown in the top plots. The spectra in the top plots are from the different substrate finishes with an SDS loading of ~2.1 µg cm^{-2}, the three spectra have had baseline subtracted in accordance with the integration method described in Section 5.3. The plots on the left are for the CH region and the plots on the right for two peaks in the fingerprint region.

This difference in intensity may be (at least partially) attributed to topographical phenomena, as can be explained by reference to Figures 6.2 and 6.10. As noted in Section 6.2 some of the solution from the aerosol droplets flows into surface scratches and grooves before solvent evaporation is complete. At the 80° grazing-angle, some of the analyte lying in the grooves and in the shadow of ridges is not exposed to the probing radiation, leading to the underestimation of surface loadings on the rougher surfaces, as shown in the predicted *versus* true loadings figures of Section 6.6 .

Changes in the spectral shapes of the CH region are more difficult to explain, but may be due to changes in molecular order in the analyte film, as discussed in Section 2.4.6. It is possible that the preparation method leads to some alignment of the SDS hydrocarbon chains, the degree of which is affected by the surface topography at a molecular level. The SEM images show that the polished surface is relatively devoid of surface texture, whereas the surfaces of smooth and rough finishes are entirely covered by scratches and

grooves (Figures 6.2 and 6.10). The less "rough" the surface finish the less disruption to the SDS film and the greater the alignment of the hydrocarbon chains. This conclusion is supported by the behaviour in the fingerprint region where the differences are only in the spectral intensities. Since most of the bands in this region are due to functional groups, such as the sulfonic head group, this region would be less affected by alignment of the CH tails, especially if the SDS adopted an arrangement where the sulfonic head group is closest to the surface.

In light of the changes of both shape and intensity of spectral features, it is not surprising that a PLS model built using spectra from all three surface finishes fails to give a reliable calibration. But it *is* surprising that PLS models built using two substrate finishes (Section 6.5) are apparently successful. So how do these calibrations work and how good are they?

One possibility is that the success of the two-finish calibrations is related to the changes in spectral shape in the CH region. Effectively, the PLS calibration uses the differences in spectral shape to determine the substrate finish from which the spectrum was collected and then determines the surface loading based on the intensity information. If this is correct, the CH spectral region should need to be included in the model, along with a sufficiently high rank to accommodate the band shape changes. The results for PLS calibrations using a combined calibration set of polished and smooth samples in the two individual spectral regions are shown in Table 6.9. It can be concluded, from larger RMSECV values, that when the CH region is unavailable the performance of the optimum model does indeed deteriorate.

Table 6.9: The selected optimum ranks and corresponding RMSECV values in μg cm^{-2} for polished and smooth combined two-finish PLS models.

background surface:		polished			smooth		
spectral region:		combined	CH	fingerprint	combined	CH	fingerprint
NA:A	rank	6	7	7	8	5	5
	RMSECV	0.131	0.143	0.186	0.15	0.168	0.266
A:A	rank	7	7	7	7	5	5
	RMSECV	0.087	0.144	0.182	0.158	0.167	0.261

The results for the combined smooth and rough sample calibrations, shown in Table 6.10, do not show the same expected trend as for the polished and smooth calibration, probably since the smooth and rough substrate finishes are microscopically more similar, as indicated by the SEM images. Hence the SDS films are more alike and the worst calibrations are, in fact, produced by using the CH region only.

Table 6.10: The selected optimum ranks and corresponding RMSECV values in μg cm^{-2} for smooth and rough combined two-finish PLS models.

background surface:		smooth			rough		
spectral region:		combined	CH	fingerprint	combined	CH	fingerprint
NA:A	rank	5	7	4	5	6	4
	RMSECV	0.122	0.169	0.132	0.143	0.217	0.158
A:A	rank	5	7	4	5	6	4
	RMSECV	0.12	0.156	0.129	0.149	0.212	0.16

These conclusions are reinforced by the results from the combined three roughness PLS calibration since it struggles to accurately predict the sample surface loadings, especially for the polished samples when only the fingerprint region is available to the PLS model and when smaller ranks are used.

6.8 Conclusions

The first conclusion from this work is that successful PLS calibrations can be prepared for SDS films prepared on stainless steel substrates where all the samples in the calibration set are prepared on the same substrate finish. This was extended to the situation where a different substrate finish was used for the background single-channel, although the best results are when the sample and background substrate finishes are the same.

Next the possibility of combined two-finish models was investigated. Results with calibration sets containing samples prepared on two different surface finishes were positive. The PLS models remained sensitive to the substrate finish used for the background single channel with smaller RMSECV values being achieved for the smoother background surface finishes.

When the model was extended to three surface finishes, significant problems began to develop. Visual inspection and univariate analysis of the spectra from the three finishes highlighted differences in both intensities and shapes. This raised the question as to how a PLS model predicts SDS loadings in a two-finish model? After some investigation it was found that the success of the two-finish combined models is due to the PLS model being able to discriminate between samples prepared on different substrates.

Given the reliance of the PLS model on subtle spectral information, a robust calibration model for samples prepared for substrates with a wide range of surface roughness is unlikely. In practical terms, this may not be a significant problem in the pharmaceutical and food industries because finishes as rough as those used as “rough” surface in this study would be unacceptable for process equipment.

6.9 References

1. Caron, J. and Jacquet, D. Infrared specular reflection spectroscopy of rough metallic substrates, *Applied Spectroscopy*, **2005**, *59*, 904-911.
2. Ogilvy, J. A. *Theory of Wave Scattering from Random Rough Surfaces*; Adam Hilger: Bristrol, 1991.
3. Hasegawa, T., Kobayashi, Y., Nishijo, J. and Umemura, J. The effect of surface roughness on infrared external reflection spectroscopy, *Vibrational Spectroscopy*, **1999**, *19*, 199-203.
4. Greenler, R. G. Design of a reflection-absorption experiment for studying IR-spectrum of molecules adsorbed on a metal-surface, *Journal of Vacuum Science & Technology*, **1975**, *12*, 1410-1417.
5. Merck Chemical Database. ChemDAT, http://www.merck.de/.
6. Martens, H. A. and Dardenne, P. Validation and verification of regression in small data sets, *Chemometrics and Intelligent Laboratory Systems*, **1998**, *44*, 99-121.
7. Kalivas, J. H. and Gemperline, P. J. In *Practical Guide to Chemometrics,* Second ed.; Gemperline, P. J.; CRC: Boca Raton, Florida, USA, 2006, 105-166.

Chapter 7

SDS and paracetamol on a glass substrate

7.1 Introduction

Previous chapters have concerned PLS calibrations built using IRRAS collected from metal surfaces. In this chapter, the calibration method is extended to spectra collected from dielectric surfaces.

The most commonly used dielectric surface in the pharmaceutical industry is glass, which has a number of desirable properties: it has good chemical resistance; it is hard and resists scratching (unlike metal surfaces where scratches can aid the build up of contamination); and can readily form smooth flat surfaces that help to prevent contamination build up. It has many applications, from small scale glassware used in testing facilities and laboratories, up to large scale glass-lined reactors.

Unlike metals, glass has strong IR features, as covered in Chapter two. This will complicate calibrations prepared using IRRA spectra since analyte spectral features that are close to glass features are likely to be swamped or distorted. Another important practical issue is the possibility that there will be more than one analyte of interest on a surface. In the pharmaceutical industry, cleaning validation may necessitate the simultaneous determination of the surface loadings of several analytes in the presence of each other. For example, in a situation where a reactor used to prepare one batch of drugs has been cleaned before being used in another process, the reactor surface may contain traces of the earlier drug and the cleaning agent. To model this situation, the work in this chapter concerns surfaces that contain trace amounts of paracetamol and SDS.

7.2 Sample preparation and characterisation

The first part of this chapter concerns quantifying SDS as a single compound on glass. Twenty-three samples were prepared by the spray method using solutions of SDS in Milli-Q water or methanol. The loading range 0–12.5 μg cm^{-2} was much broader than for the metal surfaces, mainly because these experiments were conducted early in the research programme when the target loading ranges were 0–10 μg cm^{-2} and before it became clear that surface loadings greater than 2 μg cm^{-2} were visible to the eye. As well as this, the SDS spectral features are much less intense on glass substrates, the presence of the glass feature was expected to influence the range over which a successful calibration could be achieved, and it was anticipated that nonlinear effects could be observed at higher loadings.

After preparation, seven spectra (four taken along one direction and three taken in the perpendicular direction) were collected from each sample in a manner similar to that for the metal samples. A typical spectrum is shown in Figure 7.1, where the CH region has been enlarged in the inset. The dominant feature, at ~1200 cm^{-1}, is due to glass. A major difference in comparison with the spectra from metal substrates is that the SDS features at higher wavenumbers (to the blue) of the glass feature are negative whereas those at lower wavenumbers (to the red) are positive, as discussed in Chapter two.

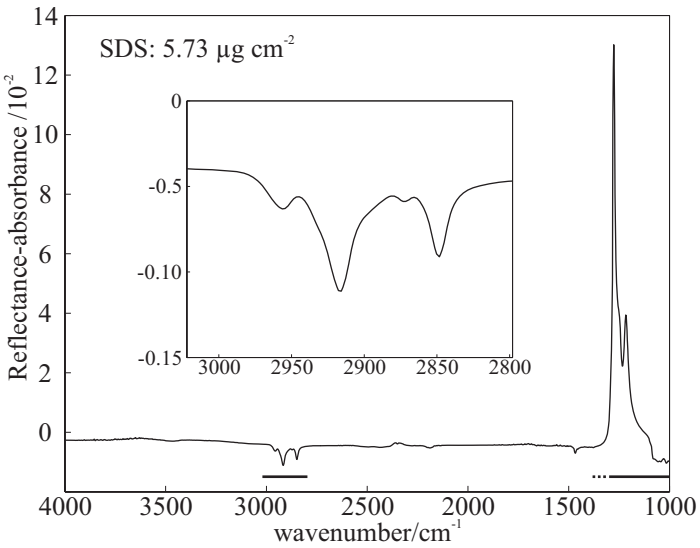


Figure 7.1: Typical IRRA spectrum for SDS at a surface loading of 5.73 µg cm^{-2} on a glass substrate. The horizontal bars above the *x*-axis indicate the SDS peaks used in the quantitative analysis.

To investigate samples with mixed analytes, thirty-six samples of SDS and paracetamol were prepared. The loading range for SDS was 0–6 µg cm^{-2} (sixteen of the SDS-only samples with loadings less the 6 µg cm^{-2} were included in the binary analyte calibration set) and for paracetamol it was 0–2 µg cm^{-2}. Since the cleaning agent (modelled by SDS) is likely to be the last analyte to come in contact with the surface, it was assumed to be present at higher concentrations than the API.

The mixtures were prepared using two variations of the spray method. The first was to spray the surface with a mixed solution of SDS and paracetamol in methanol, in which both compounds are reasonably soluble. The second, chosen to simulate the situation in the pharmaceutical industry, was to spray first with paracetamol in acetone or methanol and then with SDS in methanol or Milli-Q water. The latter method had a number of advantages over the first; the IRRA spectrum of the paracetamol could be collected before the SDS film was deposited, which meant that the calibration data set contained paracetamol-only spectra, mixed paracetamol and SDS spectra, and the SDS-only spectra from the SDS on glass calibration. Another advantage was that separate spraying allowed different solvents to be used for the two analytes.

Of the thirty-six samples, thirteen contained only paracetamol and twenty-three were mixtures of SDS and paracetamol. Ten of the mixture samples were prepared using the first spray method variation and thirteen using the second variation. Figure 7.2 shows the plot of SDS loadings *versus* paracetamol loadings.

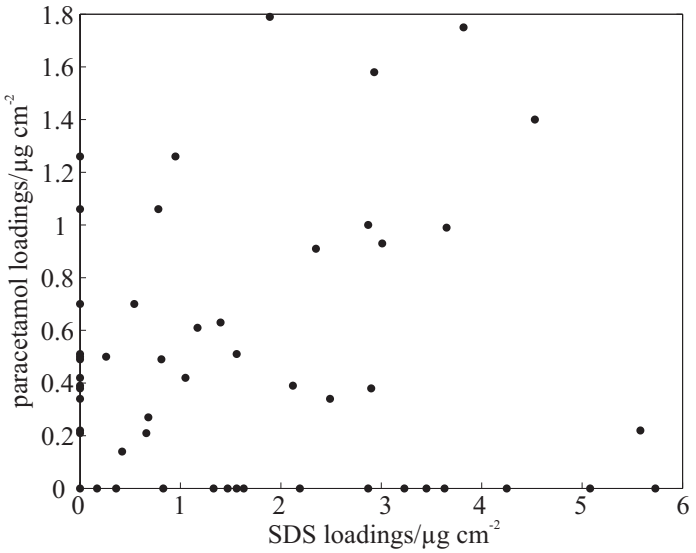


Figure 7.2: SDS surface loadings *versus* paracetamol surface loadings.

Once the samples had been prepared and left to dry, the seven spectra were collected from each. Figure 7.3 shows typical spectra for paracetamol and a mixture of paracetamol and SDS on glass substrates. The films were then rinsed from the surface into volumetric flasks. Initially, methanol was used since it dissolves both analytes, but this made it difficult to determine the SDS concentration using ^{1}H NMR (as described in Section 4.9). Later, Milli-Q water was used, which very effectively dissolves the SDS, but not paracetamol. The remaining paracetamol was rinsed from the surface using ethanol, the UV-visible spectroscopy method described in Chapter four was used to determine the paracetamol contents in both rinsing solutions, and these were summed to permit the determination of the total paracetamol surface loading.

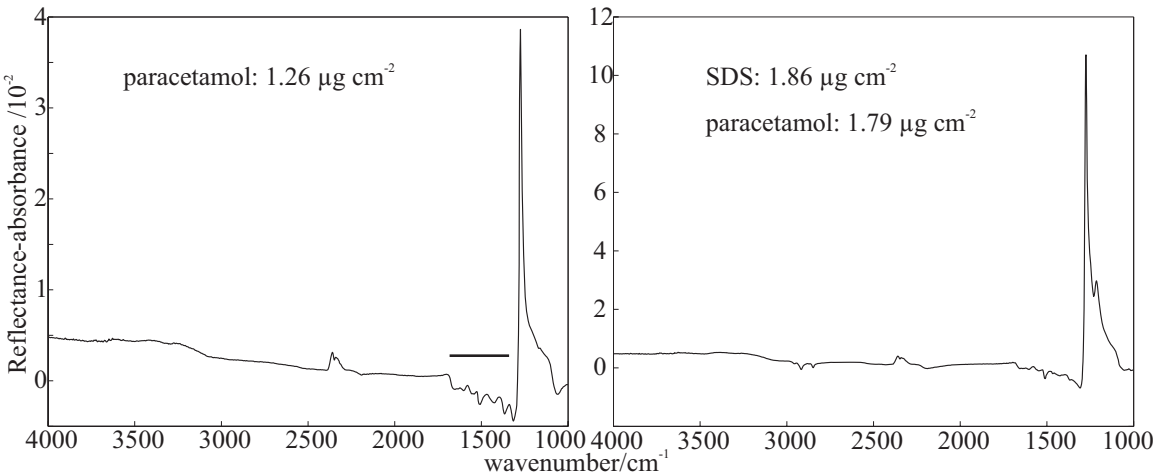


Figure 7.3: Left: IRRA spectrum for paracetamol on glass at a surface loading of 1.26 µg cm^{-2}. The horizontal bars above the *x*-axis indicate the paracetamol peaks used in the quantitative analysis. Right: IRRA spectrum for a mixture of SDS and paracetamol on glass at surface loadings of 1.86 µg cm^{-2} and 1.79 µg cm^{-2}, respectively.

Another reason for the larger SDS calibration range is related to the solvents used during sample preparation and characterisation. Solvent resonances, especially for ethanol, adversely affect the performance of the NMR calibration (especially for dilute SDS samples), since they are close to those for either or both of the SDS or/and the dioxane internal standard. This, along with baseline distortions induced by the suppression of the water NMR resonance, made it difficult to determine accurate peak integrals.

7.3 Quantitative analysis using peak area

An initial analysis of SDS films on glass was carried out using univariate statistical methods implemented using scripts written in Matlab©. Peak integrals were determined for averaged IRRA spectra from twenty-two samples using similar spectral regions as used for the metal PLS analysis. The SDS spectral features do not have the same shapes or relative intensities, but they are in the same general position. Visual inspection suggested that the performance of the integral method might be improved by inclusion of the entire glass feature, which appeared to become more intense as the SDS loading was increased. So the upper end of fingerprint region was extended from 1302 cm^{-1} to 1379 cm^{-1} (the additional region is indicated by the dashed line in Figure 7.1).

The results for the two spectral integrals, given in Figure 7.4, show strong linear relationships for SDS surface loadings. The apparently superior results achieved using the fingerprint region are due to the inclusion of the glass feature, whose intensity is directly proportional to SDS surface loading. Since the glass feature is so much more intense than the SDS bands, the integrals for the fingerprint region are less affected by baseline artefacts and noise, and hence give a better univariate calibration.

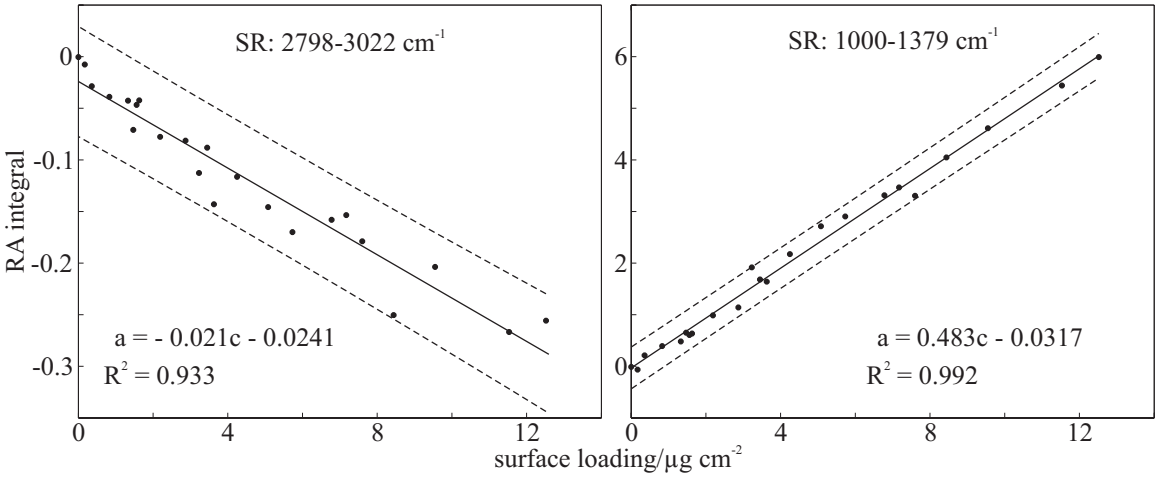


Figure 7.4: SDS band areas *versus* surface loadings for SDS on glass; in the CH region, 2798-3022 cm^{-1} (left) and the fingerprint region, 1000-1379 cm^{-1} (right). Also shown are the 95 % prediction intervals.

The results shown in Figure 7.4 suggest that a univariate integral method can produce a satisfactory calibration for the quantification of SDS on glass. However, the performance is strongly dependent on the spectral region used, and the seemingly positive effect of including the glass feature in the fingerprint range are lost in the presence of other contaminants (see further discussion in Section 7.5). It was anticipated that more robust calibrations could be achieved using multivariate quantitative analysis.

7.4 SDS on glass quantitative analysis using PLS

PLS calibrations were prepared using the same spectral regions and using scripts written in Matlab©. RMSECV *versus* rank plots calculated for the combined and individual spectral regions are shown in Figure 7.5. The optimum ranks for the different averaging approaches were selected using the methods outlined in Chapters three and six and are indicated by solid symbols.

The shapes of the RMSECV *versus* rank plots for the combined and fingerprint spectral regions, shown in Figure 7.5, are different to those seen previously. This is due to the very high intensity of the glass feature in comparison with the SDS bands. The univariate analysis indicated a strong linear relationship between the intensity of the glass feature and the SDS surface loading. This can also be recognised in the combined and fingerprint spectral regions plots since the PLS calibration improves relatively slowly as the calibration moves away from the univariate situation (a rank of one). Since the CH region does not contain any strong glass features it retains the typical RMESCV dependence on rank.

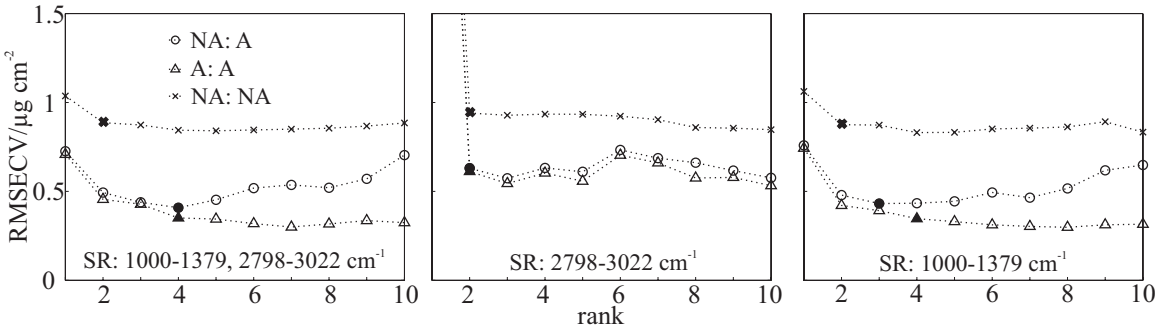


Figure 7.5: RMSECV *versus* rank plots for SDS on plates. The spectral ranges are indicated at the bottom of each plot. The three data treatment approaches are: non-averaged calibration with averaged prediction, NA:A; averaged calibration with averaged prediction, A:A; and non-averaged calibration with non-averaged prediction, NA:NA. The optimum rank for each data treatment is indicated by a solid symbol.

The optimum ranks are shown in Table 7.1, along with their corresponding RMSECV values. The RMSECV values are not directly comparable to those for the metal substrates

since the loading ranges are different. The predicted *versus* true loading plots at the optimum ranks selected in the NA:A averaging approach are shown in Figure 7.6.

7.1: The selected optimum ranks from Figure 7.5 and corresponding RMSECV values in μg cm^{-2} for PLS calibrations calculated for SDS on glass.

Averaging approach		Spectral region		
		combined	CH	fingerprint
NA:A	rank	4	2	3
	RMSECV	0.409	0.630	0.432
A:A	rank	4	2	4
	RMSECV	0.351	0.612	0.347
NA:NA	rank	2	2	2
	RMSECV	0.886	0.941	0.876

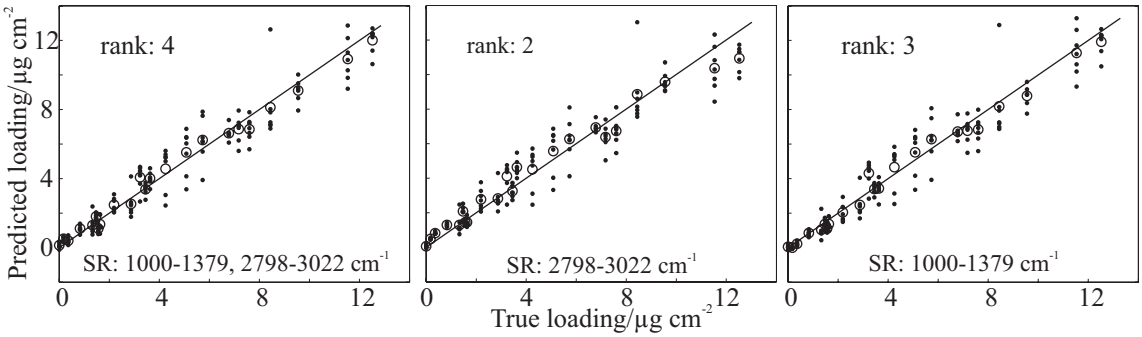


Figure 7.6: The predicted *versus* true loading plots for the NA:A models using the optimum rank shown in Figure 7.5. The dots represent the values for the non-averaged spectra and the open circles the averaged results for a single sample.

The heteroscedastic scatter of the data for each sample about their mean is readily apparent in Figure 7.6. This is mostly due to the difficulties in forming uniform films at heavier loadings, especially when the airbrush nozzle is close to the substrate surface, as required with volatile spray solutions. This hetreoscedasticity influences the method for outlier detection. If this is performed in the same manner as for the metal substrates, many samples are found to have a single outlier spectrum. The exclusion of all of these

would greatly reduce the size of the calibration set and therefore the accuracy of the calibration. To circumvent this problem, outlier detection was carried out on averaged spectra from each sample; the appropriate plots are shown in Figure 7.7.

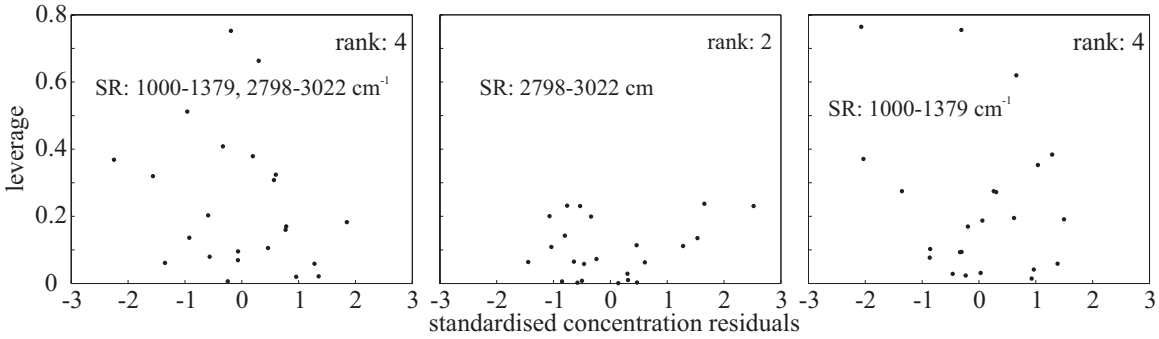


Figure 7.7: The leverage *versus* standardised concentration residuals for the spectra in the PLS calibration models using the A:A data treatment.

Figure 7.7 reveals no samples with standardised concentration residuals greater than three standard deviations. The leverages are greater than those for the metal surfaces because the averaged spectra provide fewer data points and hence each has a greater influence over the calibration. Using the $2n/m$ criterion,[1] samples considered to be outliers will have leverages greater than 27, 10 and 17 for the combined, CH, and fingerprint regions, respectively. All of the sample leverages are much less than these values and hence none were considered to be outliers.

The PLS NA:A analysis and the univariate fingerprint-region integral analysis gave similar DLs of ~0.85 and 0.83 μg cm^{-2}, respectively. This surprising result highlights the linear dependence of the glass feature intensity on concentration. Note also that the DL for the PLS analysis is a conservative estimate and, due to heteroscedasticity (following the discussion in Chapters three and five) is likely to be an overestimate. The application of this sort of calibration to an industrial cleaning validation situation is very dependent on the target contamination levels. Research by other workers into this area (Section 1.4) suggests that the DLs seen in this work would be acceptable for the purposes of cleaning validation.

7.5 Quantitative analysis of SDS and paracetamol on glass using PLS

With the knowledge that PLS calibrations can be prepared for individual analytes on glass surfaces, the calibration set was expanded to include samples that contained paracetamol and mixtures of SDS and paracetamol.

An investigation into samples that contain solely paracetamol is not presented here since successful PLS calibrations have been demonstrated elsewhere.[2, 3] The loading range over which a linear PLS model can be calculated for paracetamol is smaller than that for SDS on glass and more comparable with the ranges used in the SDS on metal studies. Univariate studies were more difficult since the usable paracetamol spectral features (Figure 7.3) are restricted to a region that is affected by both water vapour and the glass feature.

A univariate analysis of the SDS spectral regions is presented in Figure 7.8, which is to be compared to Figure 7.4 where the same type of analysis was carried out on the averaged spectra of samples containing SDS only. The analysis of the CH (2798–3022 cm^{-1}) region is very similar for both data sets, but for the fingerprint (1000–1379 cm^{-1}) region the univariate calibration for the SDS-only samples is far superior. This is due, in part, to the overlap with the paracetamol spectral features. But mostly it is due to the behaviour of the glass feature, whose intensity is related to the amount of contamination of all types on the surface. In the SDS-only analysis, it was related to the SDS concentration only, but in the analysis of the mixed calibration set, it relates to the concentrations of both SDS and paracetamol.

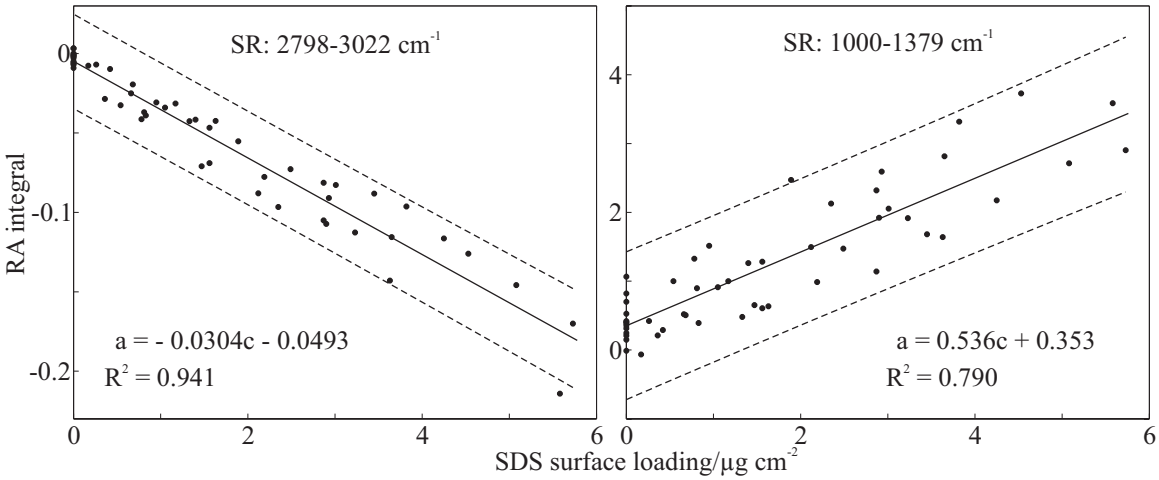


Figure 7.8: SDS surface loading *versus* peak area for the averaged spectra in the combined SDS and paracetamol calibration data set. The SDS peak integrals are calculated over the SDS features in the CH region, 2798-3022 cm^{-1} (left) and the fingerprint region, 1000-1379 cm^{-1} (right). Also shown for both regions is the 95 % prediction interval.

The linear univariate behaviour of the SDS samples may be related to the surfactant properties of SDS, which spreads out to form uniformly thick "spots" on the substrate surface. The fact that similar analyses of paracetamol-only samples have been less successful,[3] may be due to varying film thickness. The SEM images in Figure 7.9 were taken from polished stainless steel substrates, but the behaviour on glass is likely to be similar. The SDS "spots" (on the right) are circular and of relatively even thickness, whereas paracetamol spots form irregular shapes and have variable thickness

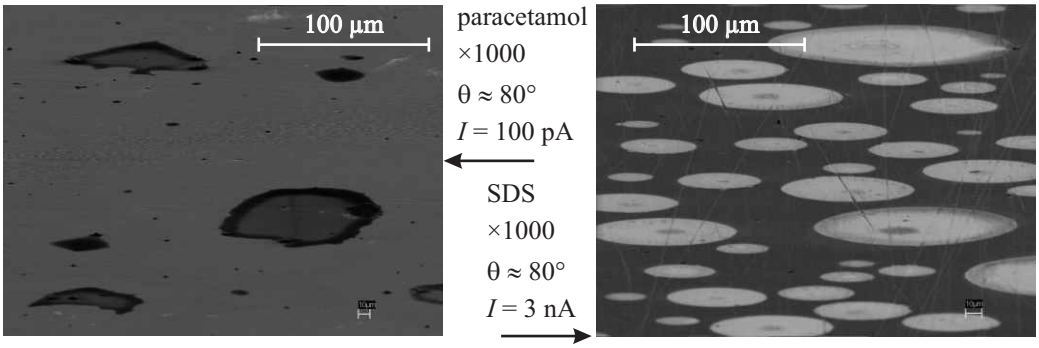


Figure 7.9: SEM images of paracetamol (left) and SDS (right) on polished stainless steel taken at ~80° to the surface normal and × 1000 magnification. The electron beam voltage is 10 kV and the electron beam currents are indicated.

Individual models for SDS and paracetamol in the presence of each other were prepared using a PLS-1 algorithm (which permits different spectral regions and model ranks to be used for each analyte) implemented in Matlab©.

Figure 7.10 contains the RMSECV *versus* rank plots for SDS in the presence of paracetamol. The shapes of these plots are very similar to those seen for the SDS-only calibration on glass, but the presence of the paracetamol and the less linear behaviour of the glass feature requires higher optimum ranks for the combined and fingerprint spectral regions. Since the CH region is unaffected by the presence of the paracetamol or glass feature, the models using this region require a rank of only two.

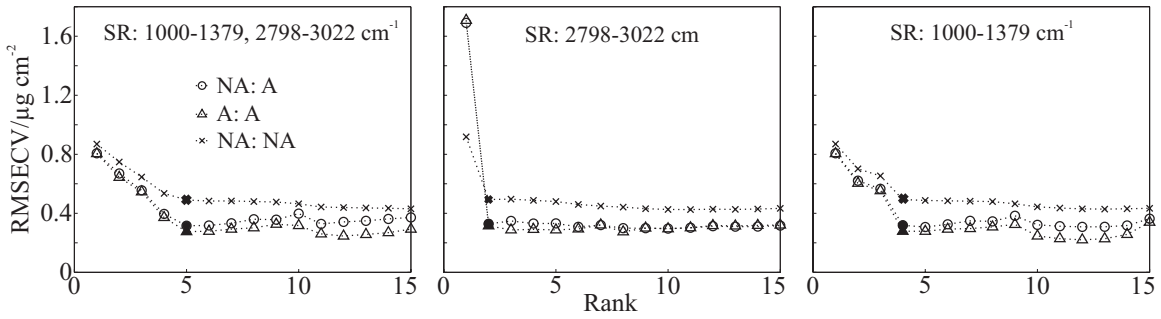


Figure 7.10: RMSECV *versus* rank plots for SDS in the presence of paracetamol on glass plates. The spectral regions are the same as used in the SDS-only analyses and are shown on the plots. The three data treatment approaches are: non-averaged calibration with averaged prediction, NA:A; averaged calibration with averaged prediction, A:A; and non-averaged calibration with non-averaged prediction, NA:NA. The optimum rank for each data treatment is indicated by a solid symbol.

The optimum ranks and corresponding RMSECV values for the PLS-1 analysis of SDS in the presence of paracetamol are given in the middle column of Table 7.2. Since they pertain to different loading ranges, the RMSECV values cannot be directly compared with those in Table 7.1. Instead, the SDS-only calibration was repeated over a reduced loading range (0–6 µg cm^{-2}) and the results are shown in the right-most column of Table 7.2. An interesting feature of this analysis in comparison with the original range (Table 7.1) is that smaller ranks are required for both the combined and fingerprint region models. This probably points to a small degree of non-linearity of the glass feature at larger (> 6 µg cm^{-2}) surface loadings.

Table 7.2: Selected optimum ranks and corresponding RMSECV values in μg cm^{-2} from Figure 7.10 compared with the results for SDS-only calibration for samples with loadings less than 6 μg cm^{-2}.

combined spectral regions (1000-1379 cm^{-1} & 2798 – 3022 cm^{-1})			
averaging approach		SDS and paracetamol	SDS only < 6 μg cm^{-2}
NA:A	rank	5	2
	RMSECV	0.316	0.356
A:A	rank	5	2
	RMSECV	0.276	0.356
NA:NA	rank	5	3
	RMSECV	0.490	0.579
CH spectral region (2798 – 3022 cm^{-1})			
averaging approach		SDS and paracetamol	SDS only < 6 μg cm^{-2}
NA:A	rank	2	2
	RMSECV	0.328	0.395
A:A	rank	2	2
	RMSECV	0.315	0.370
NA:NA	rank	2	2
	RMSECV	0.494	0.628
fingerprint spectral region (1000-1379 cm^{-1})			
averaging approach		SDS and paracetamol	SDS only < 6 μg cm^{-2}
NA:A	rank	4	2
	RMSECV	0.319	0.325
A:A	rank	4	2
	RMSECV	0.278	0.324
NA:NA	rank	4	2
	RMSECV	0.493	0.594

In general, lower RMSECV values were achieved with the combined SDS and paracetamol data set, although this required higher rank models in both the combined and fingerprint spectral regions to account for the presence of paracetamol and changes in the

glass feature. The improvement in the CH region models can be attributed to the larger number of samples in the combined data set.

The predicted *versus* true plots at the optimum ranks in the NA:A averaging approach are shown in Figure 7.11. Clearly visible in these plots is the heteroscedastic scatter of the data for each sample about their average.

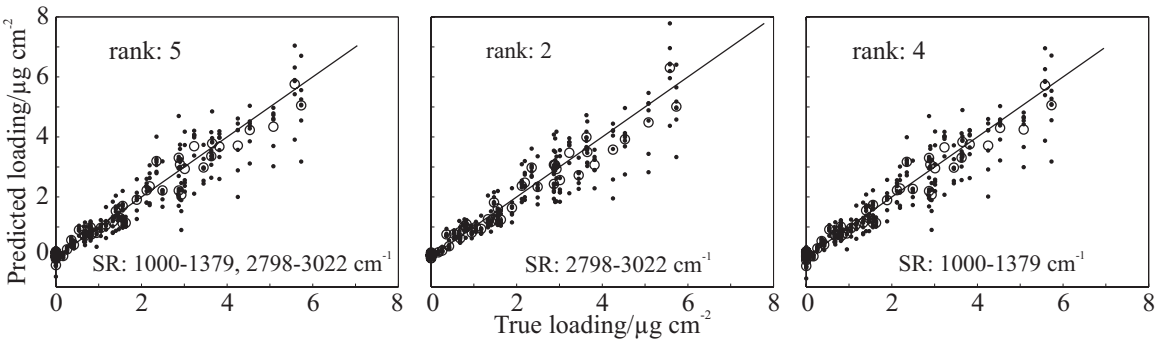


Figure 7.11: The predicted *versus* true loading plots for the NA:A models using the optimum rank shown in Figure 7.10. The dots represent the values for the non-averaged spectra and the open circles the averaged results for a single sample.

The results for paracetamol on glass in the presence of SDS are shown in Figure 7.12, using one spectral region 1684–1340 cm^{-1} (see Figure 7.3). The results gave low RMSECV values, similar to those achieved in the metal calibrations (Chapters five and six) over similar loading ranges.

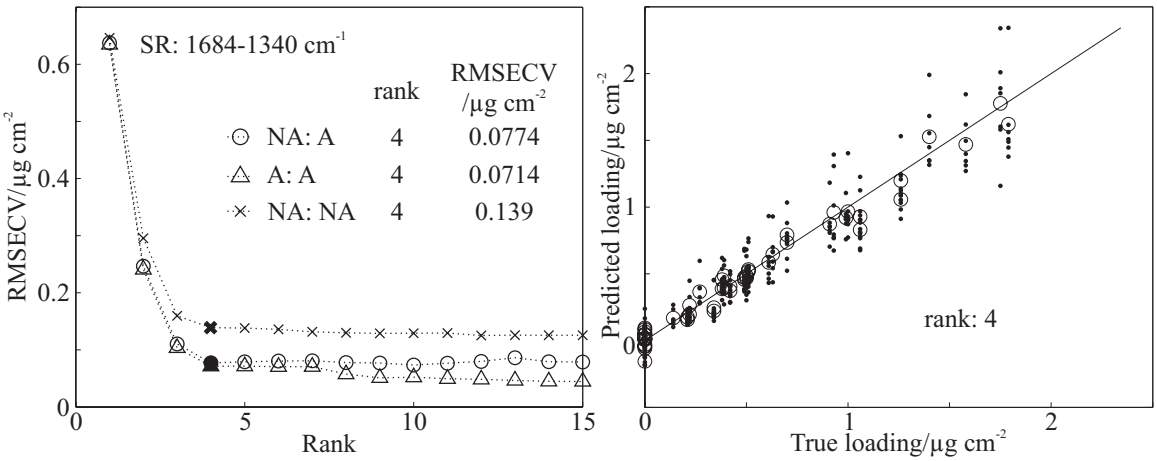


Figure 7.12: Left: RMSECV *versus* rank plots for paracetamol in the presence of SDS on glass using a single spectral region, as indicated. The three data treatment approaches are: non-averaged calibration with averaged prediction, NA:A; averaged calibration with averaged prediction, A:A; and non-averaged calibration with non-averaged prediction, NA:NA. The optimum rank for each data treatment is indicated by a solid symbol and given on the plot along with its corresponding RMSECV values. Right: the predicted *versus* true loading plots for the NA:A model using the optimum rank shown on the left. The dots are data for the non-averaged spectra and the open circles the averaged data for a single sample.

Because of heteroscedasticity, outlier detection was carried out in a manner similar to that for the SDS-only calibration, with the average data being analysed. Figure 7.13 shows that, for both analytes, there were no samples with unreasonable standardised concentration residuals, and that the leverages were all much lower than the trigger point 2*n*/*m* criterion. The sample with a relatively high leverage (~0.5) in the SDS analysis had a near-zero residual and was not considered to be a problem.

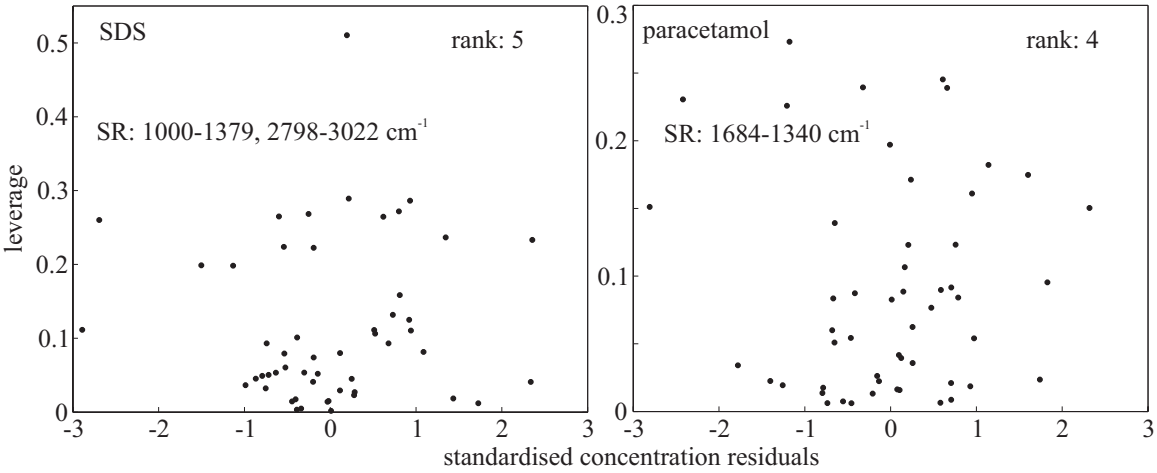


Figure 7.13: Leverage *versus* standardised concentration residuals for SDS (left) and paracetamol (right) for the optimised PLS-1 calibration models using the A:A averaging approach.

With the inclusion of paracetamol on the substrate surface, the application of an univariate integral based calibration method becomes more unfavourable for the quantification of SDS contamination. However, the multivariate PLS analysis produces similar results with and without the presence of paracetamol. This very positive result demonstrates the advantages of multivariate analysis.

7.6 Conclusions

The results in this chapter show that a linear PLS calibration can be prepared for SDS on glass over the loading range 0–12 μg cm^{-2} in the absence other contamination. Univariate calibrations, including the spectral glass feature, also work well.

PLS-1 calibrations for SDS and paracetamol in the presence of each other were investigated. The loading range for the SDS films was reduced to 0–6 μg cm^{-2}, better reflecting what might be expected in the pharmaceutical industry. Successful calibrations were achieved for SDS, although larger optimum ranks were required than for SDS alone, mainly due to the changes in the intensity of the glass feature. The loading range for paracetamol was 0–2 μg cm^{-2}, over which it has been shown (elsewhere) that successful linear PLS calibrations can be prepared in the absence of other surface species.[3] Successful calibrations were prepared for paracetamol in the presence of SDS, with similar RMSECV values as seen for the SDS on metal substrates.

This work shows that PLS modelling can be used to calibrate binary analytes (in this case, SDS and paracetamol) on glass substrates. However, caution must be exercised when investigating analyte spectral features close to the glass feature since its intensity depends on other contaminants in addition to the analytes of interest.

7.7 References

1. Kalivas, J. H. and Gemperline, P. J. In *Practical Guide to Chemometrics,* Second ed.; Gemperline, P. J.; CRC: Boca Raton, Florida, USA, 2006, 105-166.
2. Hamilton, M. L., Perston, B. B., Harland, P. W., Williamson, B. E., Thomson, M. A. and Melling, P. J. Grazing-angle fiber-optic IR reflection-absorption spectrometry (IRRAS) for in situ cleaning validation, *Organic Process Research & Development*, **2005**, *9*, 337-343.
3. Perston, B. B. Ph.D. Thesis, University of Canterbury, Christchurch, 2006.

Chapter 8

SDS on flexible dielectric substrates

8.1 Introduction

Chapter seven described PLS calibrations built using IRRA spectra collected from glass, a rigid dielectric substrate that itself has a strong IR spectral band. Interest then extended to the possibility that the IRRAS and PLS techniques could be extended to cope with spectra collected from organic films on flexible dielectric substrates, such as ethylene propylene-diene monomer (EPDM) and silicone.[1] These are polymeric materials which have IR spectra that overlap and interfere with the spectra of relevant analytes. In the preliminary study that follows, the analyte was chosen to be SDS since its preparation and spectroscopic properties had been well established in prior work.

8.2 Sample preparation and characterisation

The physical appearances of the EPDM and silicone materials are described in Chapter four. The IRRA spectra of these substrates, using a clean, polished stainless steel surface as the background single channel, are shown in Figure 8.1.

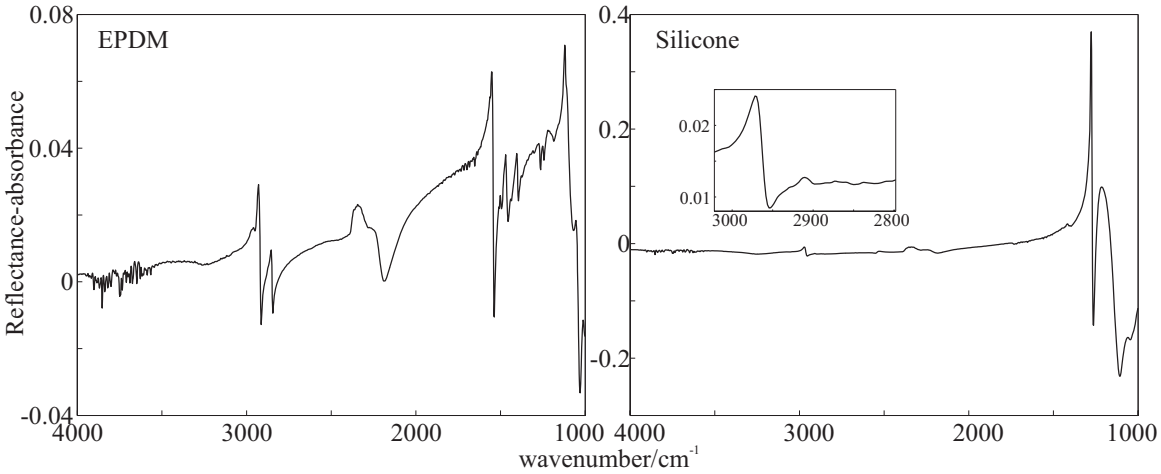


Figure 8.1: IRRAS spectra for the dielectric substrates EPDM (left) and silicone (right) using a clean, polished stainless steel surface for the background single channel.

Figure 8.1 shows that both of these materials have IR features in the regions of interest for SDS: the CH region at ~2800–3000 cm^{-1} and the fingerprint region at ~1000–1300 cm^{-1}. Other features due to atmospheric gases and the optic-fibre are also evident in the EPDM spectrum: the sharp peaks above 3500 cm^{-1} are due to water vapour; the positive feature around ~2300 cm^{-1} is due to CO_2; and the negative feature at ~2200 cm^{-1} is due to H-Se stretching vibrations of the fibre optic, (the sign of these spectral features depends on the relative conditions under which the sample and background single-channel spectra were recorded). The silicone spectrum appears to be less complex, but this is an impression influenced by the relative ordinate scales.

The favoured spray method for sample preparation could not be employed in this work, for two reasons. First, since the size of the coupons was reduced to allow them to be accommodated within a single footprint of the grazing-angle head (see Section 4.3), the SDS concentration in the rinsate was very low, even when a small volume of rinsing solvent was used; this made the primary calibration by ^1H NMR very difficult. Second, and more importantly, compounds that leached from the substrates during rinsing had ^1H NMR resonances that overlapped with the SDS signals and therefore interfered with the calculations of the peak integrals. To avoid these issues the basic smear method, described in Section 4.7, was used.

Smearing solutions were made up in either methanol or Milli-Q water. Milli-Q water was favoured as the better solvent for SDS, but has a major disadvantage that the higher surface tension makes smearing difficult. Fortunately, the slightly porous and rough finish of the polymer surfaces promoted the formation of small, separate droplets that covered the entire substrate surface and dried to leave reasonably homogeneous coverages. Volumes of ~1 mL of smearing solution, measured using an adjustable micropipette, were used for each sample. Surface loadings in the range 0–4.1 μg cm^{-2} were achieved by adjusting the dispensed volume and/or the SDS concentration of the smearing solution. Five EPDM plates were cut and cleaned, from which fifty-nine SDS samples were prepared. In the case of silicone, sixty-four SDS samples were prepared using four plates. To gauge film homogeneity and reproducibility, several samples were prepared at the same loading, at a time and using the same smearing solution, but using different plates.

Typical IRRA spectra for SDS on EPDM and silicone are shown in Figure 8.2. The horizontal lines above the *x*-axes indicate the spectral regions used in the quantitative analysis. Initial PLS quantitative analyses using the same wavenumber ranges as used in the metal and glass analyses produced poor results, which was attributed to the complexity of the substrate spectra. To overcome this problem, more appropriate wavenumber ranges were selected with the aid of a spectral region optimiser included in the OPUS chemometrics package[2] and by applying knowledge about the SDS spectrum gained during the research. For silicone, the best models were prepared when the fingerprint range was reduced to 1000–1078 cm^{-1} (to avoid the inclusion of the strong substrate features around 1200 cm^{-1}) and the CH range was increased to 2798–3022 cm^{-1} (to include all of the CH spectral features). For EPDM, the inclusion of the CH region did not improve the calibration and complex structure in the fingerprint region meant that the best results were obtained when this range was reduced to 1136–1340 cm^{-1}.

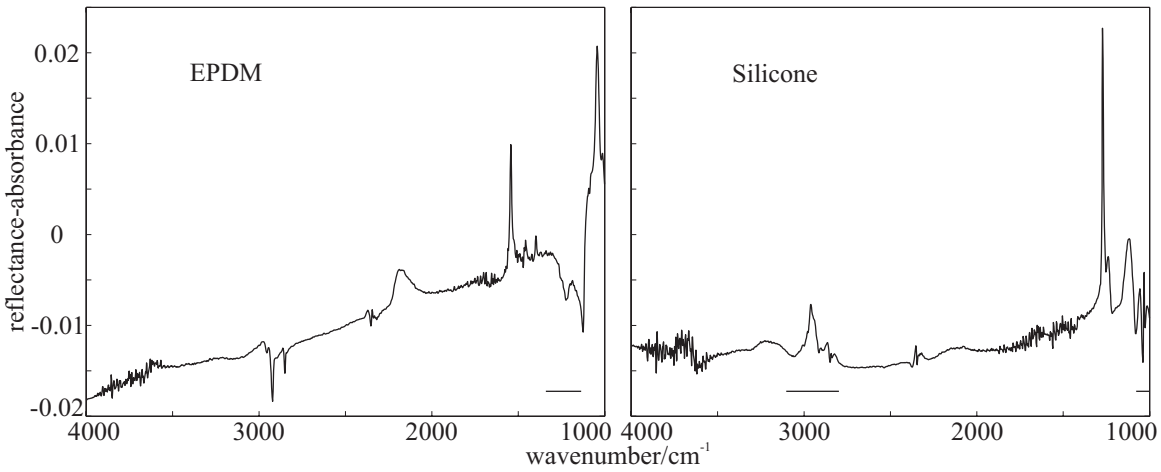


Figure 8.2: Typical IRRA spectra for SDS films on the dielectric substrates EPDM (left) and silicone (right); surface loadings are 4.10 and 4.09 μg cm^{-2}, respectively. The horizontal lines above the x-axes indicate the spectral regions used in the quantitative analysis.

8.3 Quantitative analysis

Univariate analysis was not carried out, since the substrate bands made it difficult to isolate SDS features, and the loading dependent changes in intensity of the substrate features were quite non-linear.

In this study, only one spectrum was collected from each sample and the individual spectra could therefore be treated independently using the NA:NA- type analysis referred to in earlier chapters. However, since the same background single-channel was used for all samples prepared at the same time (but on different plates), there was a possibility that the commonality of background could influence the calibration. For this reason, the NA:NA approach was modified so that all spectra that used the same background single-channel were removed together during the cross validation process. This data treatment variation is referred to as NA:NA b in the following analysis.

Figure 8.3 shows the RMSECV *versus* rank plots for EPDM and silicone PLS models; the selected optimum ranks and corresponding RMSECV values are given in Table 8.1. The shape of the RMSECV curves are similar to those seen in previous chapters, although the silicone plot has an unusual plateau in the rank range 3–5. There are small

differences between the two data treatments. Although they give the same optimum rank, the NA:NA approach gives slightly lower RMSECVs due to the inclusion of spectra with the same background in the cross validation.

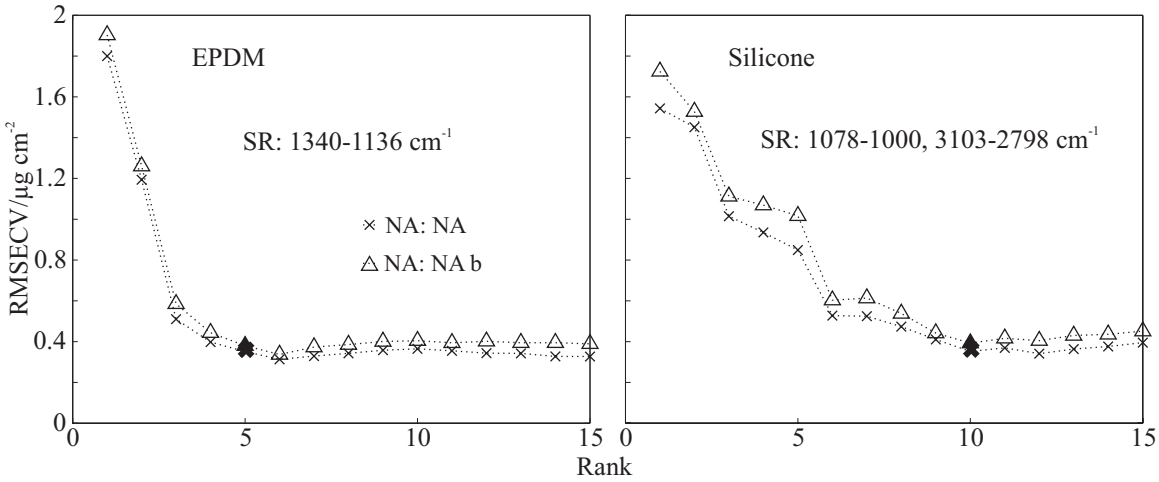


Figure 8.3: RMSECV *versus* rank plots for SDS films on EPDM (left) and silicone (right). The spectral regions are indicated in each plot. The two data treatments are: NA:NA; non-averaged calibration with averaged prediction; and NA:NA b, a modification described in the text. The optimum rank for each data series is indicated by a solid symbol.

Table 8.1: The selected optimum ranks and corresponding RMSECV values in µg cm^{-2} for PLS calibrations for SDS on EPDM and silicone.

EPDM	Spectral region: 1340–1136 cm^{-1}	NA:NA	NA:NA b
	rank	5	5
	RMSECV	0.351	0.380
silicone	Spectral region: 1078–1000, 3103–2798 cm^{-1}	NA:NA	NA:NA b
	rank	10	10
	RMSECV	0.352	0.393

The optimum ranks selected for the EPDM calibrations are similar to those for SDS-only calibrations on glass and metal substrates. This result was unexpected since the smear method produces less homogeneous samples than the spray method (see Section 4.7), and EPDM does not provide spectral windows where SDS and substrate features do not overlap. It suggests that samples were not excessively inhomogeneous and that the

calibration method can detect changes in the SDS spectrum even in the presence of an interfering substrate spectrum.

Although the surface loading ranges are different (which precludes direct comparisons) the RMSECV values at the optimum rank for SDS on EPDM and silicone are similar to those achieved for SDS on glass at surface loadings < 6 μg cm^{-2}. Better results would be achieved if more homogeneous samples could be prepared, a more accurate way of determining the surface loading could be established, and if the leaching of plasticisers from the substrate surface could be addressed.

Successful calibrations for the silicone substrate were prepared with reasonable RMSECV values but at a high optimum rank of ten. Differences from the EPDM calibrations probably result from the silicone IR spectrum, especially the strong spectral features in the fingerprint region that make it difficult for the PLS factors to describe changes in the SDS concentration. This is complicated further by the presence of the silicone bands in the CH region that interfere with the SDS CH features.

The predicted *versus* true surface loading plot for the NA:NA b data treatments using the optimum rank selected in Figure 8.3 are shown in Figure 8.4. The scatter in the data for the same true loading can be attributed to the use of the smear method for sample preparation. Even though a single spectrum is collected from each sample, the IR footprint is elliptical with an intensity profile described in Section 4.3, so the spectrum collected is still affected by inhomogeneity of the analyte films.

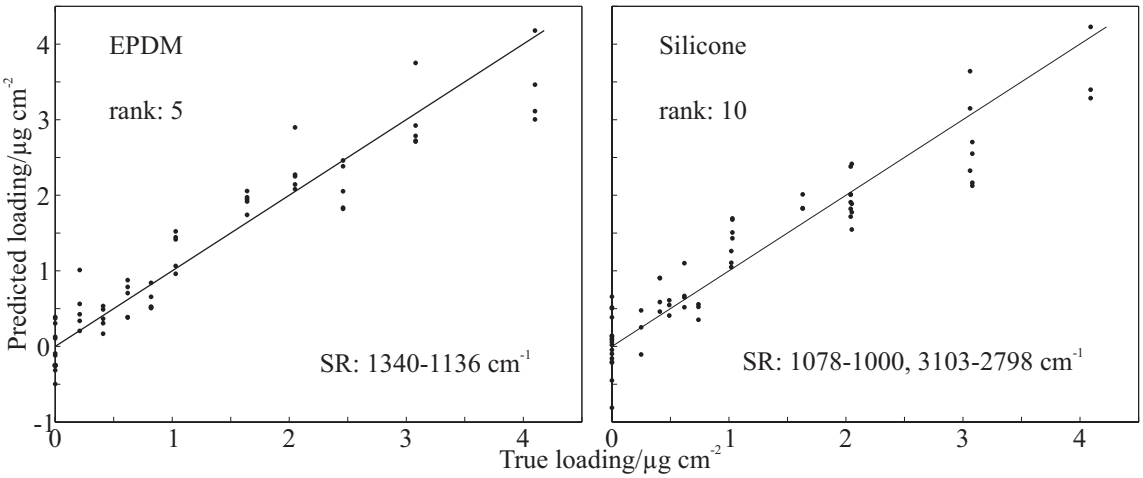


Figure 8.4: Predicted *versus* true loading plots for the NA:NA b models using the optimum rank indicated in Figure 8.3.

Outlier detection was carried out in a similar way to that described in earlier chapters. The resulting leverage *versus* standardised concentration residuals plots are shown in Figure 8.5. All of the sample leverages are well under the 2*n*/*m* criterion and none of the spectra have residuals outside the range of ± 3 standard deviations.

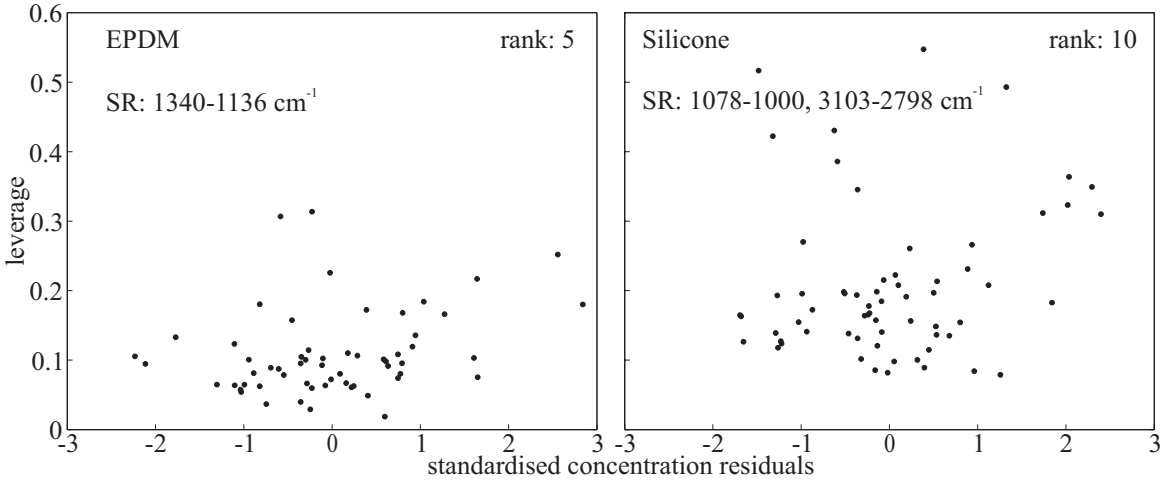


Figure 8.5: Leverage *versus* standardised concentration residuals for SDS on EPDM (left) and silicone (right) for the optimum PLS calibration models using the NA:NA b data treatment.

8.4 Conclusions

The results presented in this chapter show that linear PLS calibrations can be prepared for SDS loadings in the range 0–4.1 μg cm^{-2} on the surfaces of flexible, dielectric, polymer materials like EPDM and silicone, with the caveat that higher ranks are required. They also demonstrate that the smear method can be used to prepare samples with acceptable homogeneity to give PLS calibrations with RMSECV values similar to those obtained for SDS on glass at loadings < 6 μg cm^{-2}. Although the loading ranges are different and the RMSECV values should therefore not be directly compared, this general comparison illustrates that the PLS modelling has worked and the results are of a similar quality.

8.5 References

1. Thomson, M. and Melling, P. J.; Hamilton, M. L., 2006, Personal communication regarding: Industrial applications of the IRRAS technique.
2. Bruker Optics Inc, 4.2 Build 4, 2, 37 (20030313) ed., 1997.

Chapter 9

An industrial cleaner on a polished stainless steel substrate

9.1 Introduction

SDS was used as a model cleaning agent in the research presented in earlier chapters. It has advantages over industrial cleaning agents of being available in a pure form with known physical properties and a well-characterised IR spectrum. Industrial cleaners are mixtures, usually of a base surfactant, along with varying amounts of other additives. Quantitative, or even qualitative, information about the composition of an industrial cleaner is difficult to obtain since the manufacturers are generally unwilling to release product information that might be useful to a competitor.

The industrial cleaner investigated in this chapter is P3 cosa® PUR80 (hereafter referred to as P3). The first section of this chapter concerns the determination of compositional information using IR, HPLC, mass spectrometry and NMR. Much of this work was carried out by James Bull, a third-year graduate student on a University of Canterbury Summer Scholarship, under the supervision of the author of this thesis and Professor Peter Harland. The second section of the chapter discusses the quantitative analysis of P3 on polished stainless steel.

9.2 P3 cosa® PUR80 characterisation

The information about P3 provided by the suppliers and summarised in Section 4.4 arrived after the analyses reported here were completed.[1]

9.2.1 Infrared spectroscopy

The IR transmission spectrum of P3 was measured as described in Section 4.4 and is shown in Figure 9.1. It shows that P3 contains water, indicated by the broad band around ~3400 cm^{-1}. The strong pair of bands at ~2800 cm^{-1} are indicative of CH stretching modes and their relatively high intensity suggests a reasonably long CH backbone. The features in the fingerprint region indicate a number of unknown functional groups.

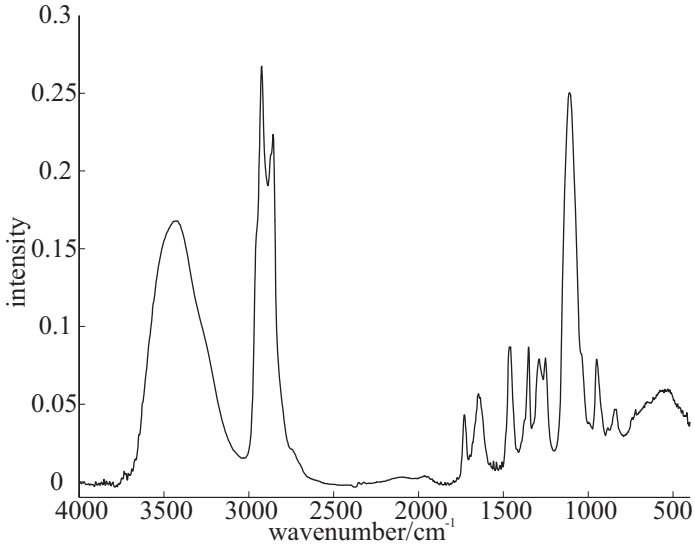


Figure 9.1: Infrared transmission spectrum for the industrial cleaner, P3-cosa® PUR80, from a thin film smeared over a KBr disk.

9.2.2 Solubility and water content

P3 is very soluble in Milli-Q water, and moisture content studies revealed that it is ~66 % water. Similar to SDS, it is reasonably soluble in methanol and only slightly soluble in ethanol, which suggests that it is surfactant based.

9.2.3 UV/vis spectroscopy

The UV/visible spectrum was measured over 200–700 nm from a 10 wt% solution of P3 in Milli-Q water, using the GBC-920 UV/visible spectrophotometer with the same parameters as used in the paracetamol analysis (2 nm resolution over a 1 cm pathlength

and a scanning speed of 180 nm min^{-1}). It is shown in Figure 9.2 and illustrates that P3 does not contain any appreciable near-UV features, apart from a weak shoulder ~280 nm on the rising absorption continuum.

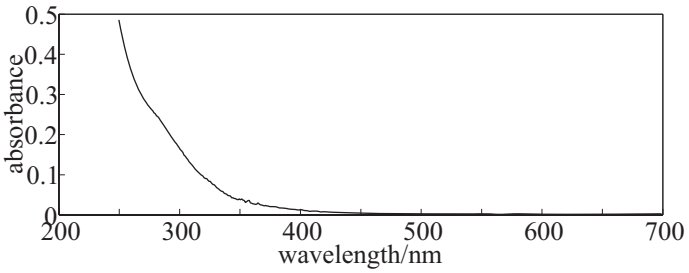


Figure 9.2: UV/visible spectrum of a 10 wt% solution P3 cosa® PUR80 in Milli-Q water.

9.2.4 HPLC and LC-MS

HPLC was carried out on a 10-fold dilution of P3 in Milli-Q water. The sample was run for 40 minutes through a C18 reverse-phase column on a standard gradient with acetonitrile to assist peak resolution. The chromatogram, obtained using a refractive index detector (RID), is shown in Figure 9.3.

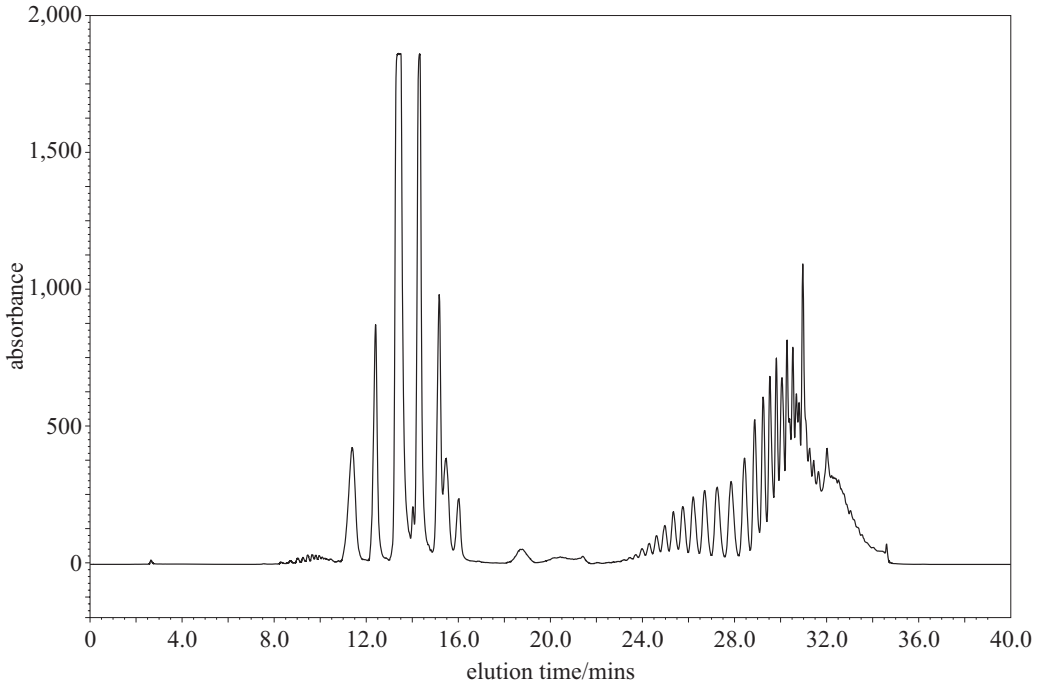


Figure 9.3: HPLC chromatogram for P3 cosa® PUR80 using a refractive index detector.

The shorter-time (more polar) manifold, eluting over the period ~11–16 minutes, probably represents a series of water soluble polymeric chains with various functional group modifications. The nature of the components in the longer time (less polar) manifold, eluting over ~23–35 minutes, is unknown.

LC-MS was carried out on a 1 mg mL^{-1} sample of P3 in Milli-Q water. Liquid chromatography, with the same gradient as used with HPLC, was used to separate the components and identification was achieved by time-of-flight mass spectrometry with electrospray ionization. This gave a total ion current trace in agreement with the HPLC trace. Selected ion analyses were carried out for all major eluting species on the total ion current trace. Analysis of this data showed that the components of the short time manifold have mass differences of 44 amu, strongly suggesting a polyethylene glycol (PEG) polymer base with $–O–CH_2–CH_2–$ subunits. The fragmentation patterns show a distribution of chain length from six to thirty PEG units with an average of ten or eleven. The presence of PEG was confirmed by a MS database search and comparison carried out by the Bruce Clark, the Department of Chemistry's mass spectrometry technician. The LC-MS study also showed that some of the PEG chains have nitrogen (most likely amine) functionality, and contain oxygen and methylene groups. The second manifold in Figure 9.3, eluting from ~23–35 minutes is more problematic, but probably relates to higher mass PEG chains with more complex functionality.

9.3 Sample preparation

Since P3 is a mixture of compounds, quantitative analysis is difficult. The "concentration" can not be defined in terms of molarity; instead it needs to be defined in terms of either weight per unit volume or volume per unit volume. Since the cleaner is reasonably viscous, with a reported density between 0.99 – 1.04 g cm^{-3}, it was convenient to use units of μg mL^{-1}.

Initially, it was intended that the spray method would be used to prepare the P3 surface samples. This required the establishment of a primary calibration method. Like SDS, P3 does not display sufficient UV absorptivity for quantitative analysis, so a modification of the quantitative ^{1}H NMR method was adopted. The ^{1}H NMR spectrum of P3 in D_2O is shown in Figure 9.4; it contains two broad resonances at ~3.6 and 1.2 ppm, plus a number of weaker resonances due to minor components. During this investigation, it was found that dilute solutions of P3 in Milli-Q water degrade over a period of days, so analyses were performed on freshly prepared solutions.

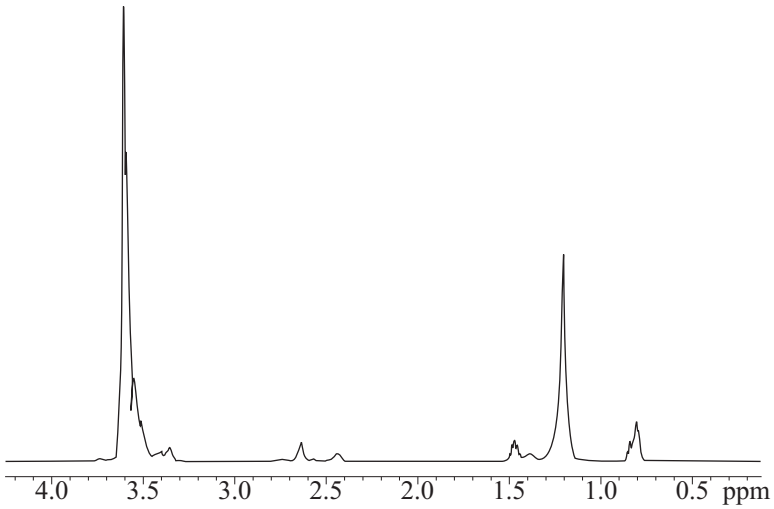


Figure 9.4: ^{1}H NMR of P3 cosa® PUR80 in D_2O.

Quantitative analysis of P3 was attempted using the ~3.6 ppm peak and the conditions described in Section 4.9. The dioxane resonance was too close to the ~3.6 ppm P3 resonance, so the internal standard was changed to acetic acid, which exhibits a NMR resonance at 2.05 ppm. Standard solutions of P3 were prepared in Milli-Q water over the concentration range 0.5–11.24 μg mL^{-1} (corresponding to surface loadings of 0.22–5.00 μg cm^{-2} using 100 mL of rinse solution). It proved to be impossible to obtain reproducible calibration curves, even with freshly prepared standards, and the ^{1}H NMR primary calibration method was eventually abandoned.

In the absence of a suitable calibration method, samples were prepared using the smear method. 600 μL aliquots of P3 dissolved in Milli-Q water were smeared over the surfaces

of 150 × 150 mm^2 polished stainless steel plates to produce thirty-nine samples with a loading in the range 0–5 μg cm^{-2} (although only six had loadings greater than 2.5 μg cm^{-2}). After drying, loadings above ~0.5 μg cm^{-2} were visible as a slightly bluish discolouration of the surface.

9.4 P3 cosa® PUR80 quantitative analysis

A typical IRRA spectrum of P3 on polished stainless steel is shown in Figure 9.5. The regions of the spectrum used in the quantitative analysis are indicated by horizontal lines above the *x*-axis. The wavenumber range for the CH region (2798–3103 cm^{-1}) is the same as used in the SDS analyses. The structure in the fingerprint region is quite different from the SDS spectrum and the range of 1000–1205 cm^{-1} was chosen to encompass the strongest analyte band in this region. Inclusion of weaker features between 1200–1700 cm^{-1} did not improve the calibration, probably due to adverse effects of baseline distortions and interferences by atmospheric water vapour.

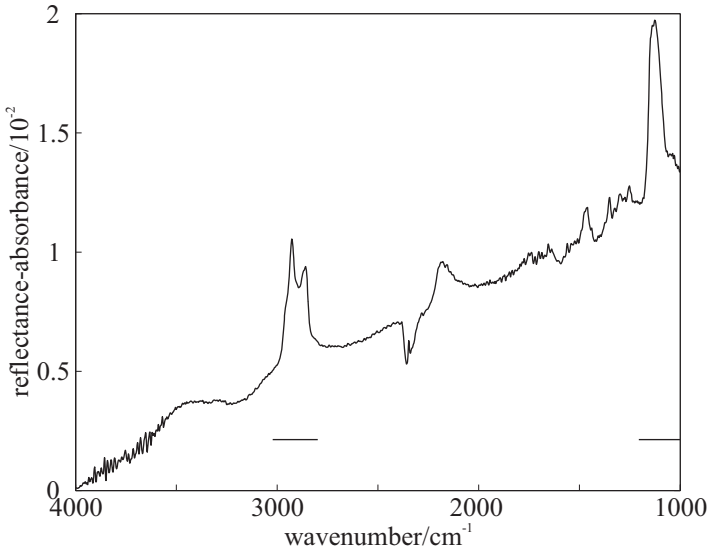


Figure 9.5: IRRA spectrum for P3 cosa® PUR80 on polished stainless steel, surface loading ~4 μg cm^{-2}. The horizontal lines above the *x*-axis indicate the spectral regions used in the quantitative analysis.

9.4.1 Univariate analysis

Univariate integral analysis was carried out on the two spectral regions using samples in the loading range, 0–2.5 μg cm^{-2} (since there were few samples with greater loadings). The resulting calibration plots, along with their 95 % prediction intervals, are shown in Figure 9.6. Since the smear method of sample preparation was employed, the averaged spectra were used to mitigate the affects of sample inhomogeneity.

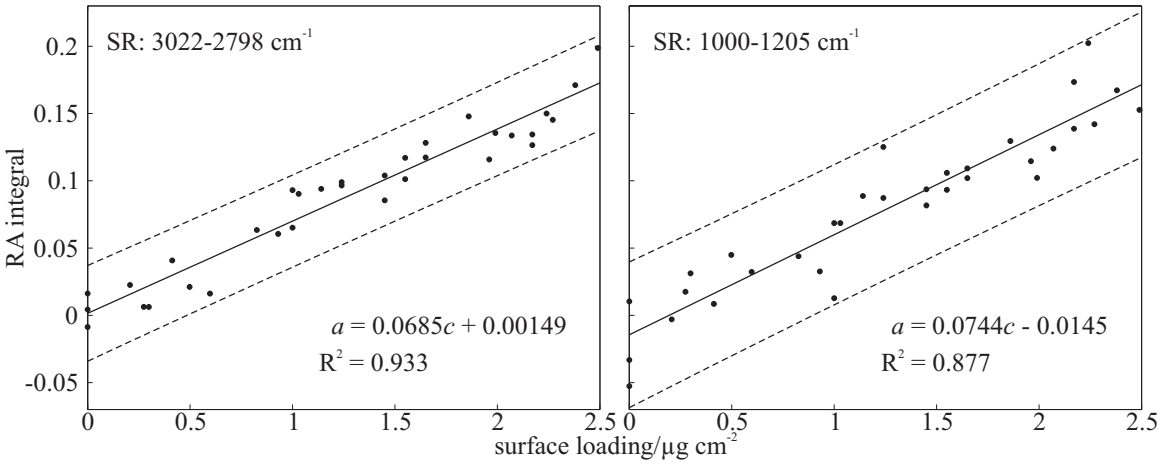


Figure 9.6: Spectral integrals for averaged spectra *versus* surface loadings for the spectral regions shown in Figure 9.5. The plot on the left is for the CH region and that on the right is for the fingerprint region. Also shown as dashed curves are the 95 % prediction intervals.

The univariate calibration for the CH region is better than that for the fingerprint region, but there is still a large prediction error as indicated by the 95% prediction intervals. The calibration uncertainty for the CH region amounts to ~ ± 0.5 μg cm^{-2}.

9.4.2 Multivariate analysis

Multivariate PLS analysis was carried out using the combined spectral regions shown in Figure 9.5. The RMSECV *versus* rank plot is shown in Figure 9.7. The same data set used in the univariate analysis was used in this analysis where the samples > 2.5 μg cm^{-2} were omitted.

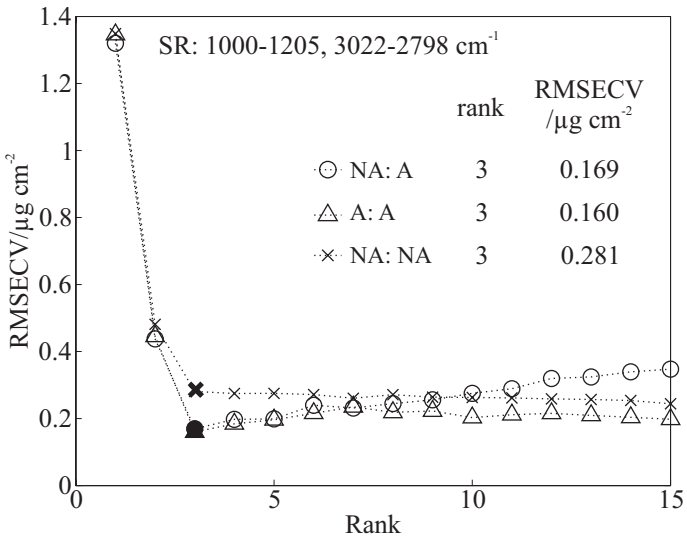


Figure 9.7: RMSECV *versus* rank plots for P3 cosa® PUR80 on polished stainless steel. The three data treatment approaches are: non-averaged calibration with averaged prediction, NA:A; averaged calibration with averaged prediction, A:A; and non-averaged calibration with non-averaged prediction, NA:NA. The optimum rank for each data treatment is indicated by a solid symbol and is recorded on the plot, along with its corresponding RMSECV value.

The shape of the plots and the values for the optimum ranks shown in Figure 9.7 are similar to those for SDS films on metal and glass substrates. This is encouraging since it suggests that, even though P3 is not a pure compound, the IRRAS and PLS techniques can still quantify surface contamination. The results for SDS (0–2.1 µg cm^{-2}) and P3 (0–2.5 µg cm^{-2}) on polished stainless steel are compared in Table 9.1. The SDS results are obviously better, with RMSECVs that are consistently ~40–45% of those for P3, even when the smaller loading range is considered. The preparation method is likely to influence these results, the SDS films being prepared by the spray method producing more homogeneous films than the smear method used for P3.

Table 9.1: A comparison of the results from SDS and P3 cosa® PUR80 on polished stainless steel.

	averaging approach:	NA:A	A:A	NA:NA
SDS	rank	**3**	**4**	**4**
	RMSECV/μg cm^{-2}	0.0754	0.0643	0.111
P3 cosa® PUR80	rank	**3**	**3**	**3**
	RMSECV/μg cm^{-2}	0.169	0.160	0.281

The resulting predicted *versus* true loading plot for the optimum model selected for the NA:A data treatment is shown in Figure 9.8. The heteroscedastic nature of the P3 uncertainty is greater than for the SDS samples on the same substrate, probably (as noted earlier) due to the inferiority of the smear sample preparation method.

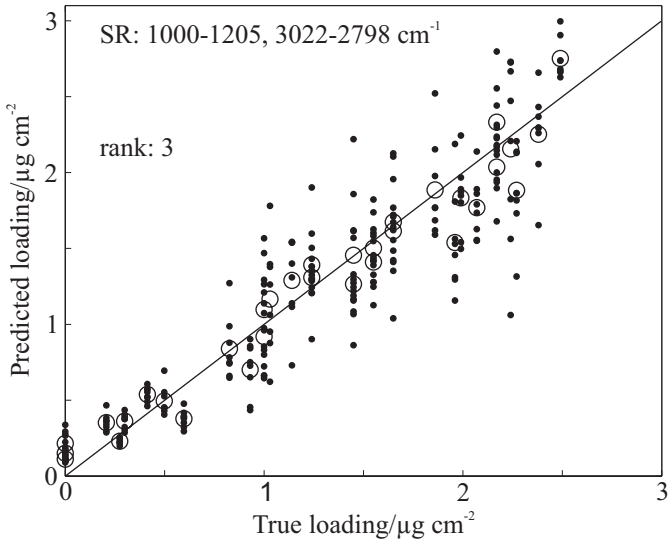


Figure 9.8: Predicted *versus* true loading plots for the NA:NA b models using the optimum rank shown in Figure 9.7.

A plot of leverages *versus* standardised concentration residuals is shown for the P3 data set in Figure 9.9. Three suspect spectra (indicated by broken circles in Figure 9.9) were detected from three samples whose other spectra had good leverages and standardised residuals. The statistical properties of these outliers did not necessitate their removal from

the data set: the two samples with large standardised residuals had small leverages, whereas the opposite was true for the third outlier.

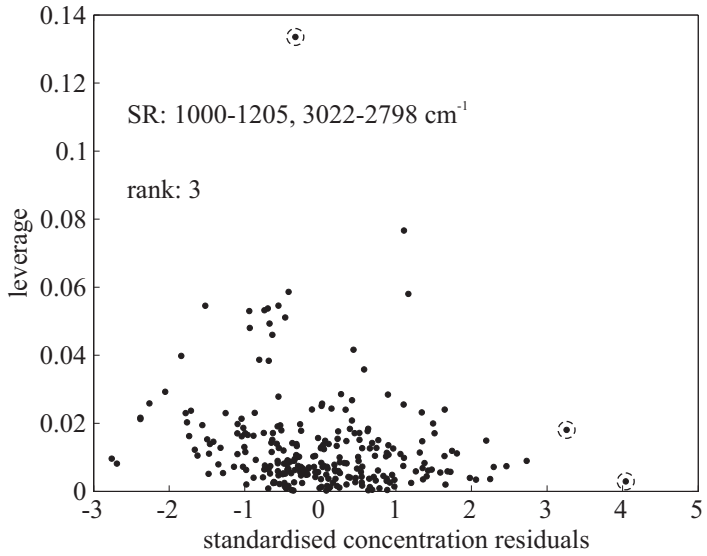


Figure 9.9: Leverage *versus* standardised concentration residuals for P3 cosa® PUR80 on polished stainless steel for the optimum PLS calibration model using the NA:NA b data treatment. The points enclosed in dashed circles are suspect samples, which required further analysis.

9.5 Conclusions

The properties of an industrial cleaner, P3 cosa® PUR80, were investigated. Initially very little compositional information was available. Through the use of IR, UV, HPLC and LC-MS it was possible to determine that P3 cosa® PUR80 has a surfactant base composed of modified PEG chains.

Using the smear method samples of P3 cosa® PUR80 were prepared on polished stainless steel substrates and PLS calibrations were generated over the surface loading range 0–2.5 μg cm^{-2}.

There are two factors associated with this work which makes it important for industrial applications. Firstly, P3 cosa® PUR80 comprises several compounds in inexactly known

amounts. Secondly, successful calibrations could be generated even when the exact composition of the analyte was unknown.

9.6 References

1. Ecolab, Cleaning Validation Handbook CD-Rom, Pharmocos, 2005.

Chapter 10

General conclusions and future directions

10.1 General conclusions

From the research presented in this thesis, and in related publications,[1-3] it has been established that IRRA spectra can be used to prepare multivariate calibrations for the quantification of trace surfactants and pharmaceuticals on glass, metal and polymer surfaces. This research could have important consequences in the pharmaceutical industry where IRRAS could be used in the detection, identification and quantification of surface contamination.

The compound of major interest in this research was SDS, which was used to model an industrial cleaning agent. PLS models were constructed for SDS on metal and dielectric surfaces and it was found that linear calibrations could be achieved for surface loadings up to 12.5 µg cm^{-2}. Since SDS surface loadings greater than ~2.5 µg cm^{-2} on metals or ~6 µg cm^{-2} on dielectrics are visible to the eye, calibrations were built with these loadings as the upper loading limit.

Successful PLS-1 calibrations were prepared for SDS and paracetamol on glass in the presence of each other. This is a very positive outcome since it is likely that real-life cleaning validation situations will require determination and quantification of multiple contaminants.

Since flexible dielectric surfaces are used in industrial environments, EPDM and silicone were also investigated. Even with difficulties in sample preparation and interference from substrate IR spectral features, useful PLS calibrations were generated for SDS loadings below ~4.1 µg cm^{-2}.

Although the preparation methods did not produce perfectly homogeneous samples, successful calibrations were achieved. In general, univariate statistical methods could be used when SDS was the only analyte present on the substrate surface. But, particularly in situations where there were multiple analytes or where the substrate had interfering spectral features, better and more robust calibrations were achieved using multivariate calibration methods.

The final section of work presented in this thesis concerned a common industrial cleaner, P3 cosa® PUR80. It was found to be a PEG based surfactant with functional group modifications, the exact nature of which are currently unknown. It also contains other unknown additives. Even in the absence of exact compositional information, it was possible to prepare a quantitative PLS calibration using IRRA spectra from samples with surface loadings < 2.5 μg cm^{-2}. This is another very positive result since it indicates that that the IRRAS technique can be used to quantify (by mass) a substance that is made up from a number of compounds.

RMSECV values for the calibration sets using the NA:A data averaging approach were < 0.41 μg cm^{-2}, which corresponds to conservative detection limit for surface residues of ~0.86 μg cm^{-2} or better. The RMSECV values are dependent on the loading range of the calibration; hence, for smaller loading ranges (such as for SDS on metal surfaces) detection limits were ~0.2 μg cm^{-2}. In Chapter one it was suggested that visual detection limits for surface contamination are between 1–4 μg cm^{-2} and this is in the order of the levels required for cleaning validation. In this work, a pessimistic approach has been adopted for the determination of detection limits for the PLS calibrations. Despite this, the calculated detection limits are lower than validation requirements. The IRRAS techniques developed in this thesis are not as sensitive as (for example) HPLC but have the major advantage of permitting *in situ* remote sensing in real-time and with a very short turn-around time. A disadvantage of HPLC techniques is that they are indirect (see Chapter one); that is, they can consider only those contaminants introduced into the chromatograph, but not any that were on the surface but were not transferred on to the

swab or into the rinse sampling. The IRRAS technique measures the contamination directly on the surface, which can lead to more accurate results on a shorter time scale. Depending on the number of scans, an *in situ* spectroscopic measurement can be made in ~15 seconds using a conventional laboratory spectrometer. With the use of an appropriate predetermined chemometric calibration, the quantification process can be completed in less than a minute, which is far quicker than wet chemical methods.

10.2 Future directions

The initial objectives and aims of the project were successfully achieved and, in many cases, the results surpassed expectations. There are a number of areas that could benefit from future research, some of which are outlined below.

10.2.1 Error analysis

The study presented in this thesis demonstrates the feasibility of IRRAS for the quantification of surface contamination, with the caveat that the calculated detection limits and confidence intervals are subject to bias. This bias is mainly due to errors in the reference concentrations, making it difficult to obtain good estimates of the sensitivity of PLS calibrations. Future research should investigate ways to better assess the sensitivity of a calibration, either through theory or experiment. With modifications to current techniques, improvements in the uncertainty associated with the primary calibration could possibly be reduced and better characterised. The preparation of truly homogeneous samples would reduce concentration errors and would improve the heteroscedastic error component seen in the current results. This would lead to a better estimate of the sensitivity in the IRRAS technique. In the current research a conservative approach to calibration sensitivity has been adopted.

10.2.2 IRRA spectra with an adjustable grazing angle

The IRRA spectra in this research were collected by using a grazing-angle head with a fixed incident angle of 80°. In Chapter two, the effect of adjusting the incidence angle was discussed; large incidence angles (like 80°) work well for metals, where there is a compromise between the maximum spectral intensity and the size of the IR footprint. The situation for dielectric substrates is not so simple. As discussed in Chapter two there may be some improvement to be gained at smaller incidence angles (although this might be at the cost of the signal-to-noise ratio).

During the course of this research an instrument was designed and built that allowed the incidence grazing-angle to be adjusted. Unfortunately, due to constraints of time and instrument availability, experiments were not conducted using this instrument. However, some work has been presented elsewhere for paracetamol films and for electrostatically self-assembled multilayers on glass.[4] Future research could focus on the effect of adjusting the incidence angle for incoming IR radiation for systems with SDS on dielectric substrates, such as glass, to establish whether better calibrations can be achieved. The theory of incidence angle effects could also be investigated in more detail with possible emphasis on analyte signal and signal-to-noise ratio.

10.2.3 Surface roughness

An investigation, presented in Chapter six, into the effects of surface roughness for stainless steel substrates resulted in some interesting conclusions: for example, an attempt to produce a robust, combined model for all three of the surface finishes investigated was unsuccessful. This was attributed, at least in part, to significant differences between the spectra collected from the three finishes. It is likely that more success could be achieved if the variations in surface finish were not as marked. Future work could focus on substrates with more similar surface finishes, with an aim to gain insight into how much variation can be accommodated by the PLS modelling process. This research would require an appropriate way of quantifying surface roughness, which could then be used to

aid in the interpretation of changes in analyte spectral features with changes in the surface finish.

Changes in the slope of the IRRA spectra baselines are often related to changes in surface roughness (see spectra and discussion in Section 6.4). It is possible that future research could more thoroughly investigate this phenomenon and exploit it as a method for quantifying substrate roughness.

10.2.4 Mixtures of analytes

Chapter seven describes experiments to quantify two coexisting analytes on glass. The two analytes investigated, SDS and paracetamol, have significantly different spectra. A successful investigation into two spectrally similar analytes, paracetamol and aspirin, has been presented elsewhere,[3, 4] but there are no studies into three or more coexisting analytes. The results presented in this thesis suggest that it should be possible, although the accuracy of the results will depend on the analytes, the errors in their surface loadings and on the substrate surface. It is postulated here that the results will not be as good as those for individual analytes, but since univariate methods are unlikely to work at all, PLS models may be acceptable in an industrial context.

10.2.5 P3 cosa® PUR80

Chapter ten covers quantitative analyses for films of the industrial cleaner, P3 cosa® PUR80, on polished stainless steel. This research focused on the elucidation of some structural information and application of PLS analysis to P3 cosa® PUR80 films prepared using the smear method. It is possible that better results could be achieved with a PLS calibration if more homogeneous films were prepared. This has been shown to be possible if the spray sample preparation method is used, but a primary calibration method is required to determine the P3 cosa® PUR80 concentration in the sample washings. Methods recommended by the manufacturer are given in Appendix B.

10.2.6 Instrument design

One of the major drawbacks to the industrial implementation of the IRRAS technique is related to the flat-base design of the grazing-angle head used in this work. Spectra can only be collected from flat surfaces of sufficient size to accommodate the entire base of the grazing-angle head. As discussed in Chapter one, cleaning validation of hard-to-clean and hard-to-reach places within a system is very important. With the current design, this could not be achieved. Future work into the miniaturisation and testing of sampling heads for hard-to-reach locations would represent a major contribution to the practical application of the IRRAS method. This could possibly be achieved using IR diode lasers to get more intense radiation at pre-selected wavelengths, making it possible to decrease the size of the IR footprint. Another area that could be investigated is adaptation of the current sampling head to cope with curved surfaces and corners that are likely to be found within a pharmaceutical reactor.

10.2.7 Alternative applications

The research reported in this thesis has been focused on possible applications in the pharmaceutical industry; this is reflected in the choice of analytes and substrates. There are many other situations that could benefit from the application of the IRRAS technique. This research has shown the quantitative calibrations can be prepared for analyte films on very different substrates. Future research could investigate different analytes with differing properties to those presented in this work; examples could include paints, explosive residues, or even biological materials, such as proteins and bacteria.

10.3 References

1. Hamilton, M. L., Perston, B. B., Harland, P. W., Williamson, B. E., Thomson, M. A. and Melling, P. J. Grazing-angle fiber-optic IR reflection-absorption spectrometry (IRRAS) for in situ cleaning validation, *Organic Process Research & Development*, **2005**, *9*, 337-343.
2. Hamilton, M. L., Perston, B. B., Harland, P. W., Williamson, B. E., Thomson, M. A. and Melling, P. J. Fiber-optic infrared reflection absorption spectroscopy for trace analysis on surfaces of varying roughness: sodium dodecyl sulfate on stainless steel, *Applied Spectroscopy*, **2006**, *60*, 516-520.
3. Perston, B. B., Hamilton, M. L., Harland, P. W., Williamson, B. E., Melling, P. J. and Thomson, M. Grazing-angle fiber-optic Fourier transform infrared reflection-absorption spectroscopy for the *in situ* detection and quantification of two active pharmaceutical ingredients on glass, *Analytical Chemistry*, **2007**, *In press*.
4. Perston, B. B. Ph.D. Thesis, University of Canterbury, Christchurch, 2006.

Appendix A

Statistics

A.1 Least squares linear regression

For a data set $\boldsymbol{c}$ and $\boldsymbol{a}$, $\boldsymbol{c}$ is a vector of the independent variables, usually the concentration or amount of an analyte, and $\boldsymbol{a}$ is a vector of the corresponding dependent variables, the output quantities from the $\boldsymbol{c}$ inputs, commonly signal intensities. The relationship between the $\boldsymbol{c}$ and $\boldsymbol{a}$ variables can be described in terms of equation (A.1), where $\boldsymbol{a}$ is described by a calibration function F($\boldsymbol{c}$) plus an error term $\mathbf{e}_{\boldsymbol{a}}$ which describes the deviations from model values.[1, 2]

$$\boldsymbol{a} = F(\boldsymbol{c}) + \mathbf{e}_{\boldsymbol{a}} \quad \text{(A.1)}$$

For a linear least squares calibration the calibration function is described in equation (A.2), which is an equation for a straight line where B_0 is the *y*-intercept and B is the slope.

$$\boldsymbol{a} = B_0 + B\boldsymbol{c} + \mathbf{e}_{\boldsymbol{a}} \quad \text{(A.2)}$$

Equation (A.2) can be modelled by the least squares estimate as shown in equation (A.3), where $\hat{\boldsymbol{a}}$, $\hat{B}_0$ and $\hat{B}$ are the least squares estimates for the independent measured variables, the y-intercept and the slope, respectively. This model holds as long as there is a linear relationship between $\boldsymbol{a}$ and $\boldsymbol{c}$ and the errors associated with these variables have a mean of zero and are uncorrelated.

$$\hat{\boldsymbol{a}} = \hat{B}_0 + \hat{B}\boldsymbol{c} \quad \text{(A.3)}$$

Residuals for the $\boldsymbol{a}$ variables (d_a) can be calculated using equation (A.4) and can be used to check the assumptions above.

$$d_a = \boldsymbol{a} - \hat{\boldsymbol{a}}$$

$$= \boldsymbol{a} - \hat{\mathrm{B}}_0 + \hat{\mathrm{B}}\boldsymbol{c} \tag{A.4}$$

Variables B_0 and B can be estimated using least squares regression on $\boldsymbol{c}$ using equations (A.5) and (A.6); this is for the homoscedastic errors. In the case of heteroscedastic errors, weighted least squares can be used, although this was not used in this work and its treatment can be found elsewhere.[1]

$$\hat{\mathrm{B}}_c = \frac{\mathrm{Q}_{ca}}{\mathrm{Q}_{cc}} \tag{A.5}$$

$$\hat{\mathrm{B}}_{0c} = \frac{\sum \boldsymbol{a} - \hat{\mathrm{B}}_c \sum \boldsymbol{c}}{m} \tag{A.6}$$

The Q parameters used in equation (A.5) can be calculated using the equations below, where $\bar{c}$ and $\bar{a}$ are the mean for the $\boldsymbol{c}$ and $\boldsymbol{a}$ variables and m is the number of samples.

$$\mathrm{Q}_{cc} = \sum (\boldsymbol{c} - \bar{c})^2$$

$$= \sum \boldsymbol{c}^2 - \frac{\left(\sum \boldsymbol{c}\right)^2}{m} \tag{A.7}$$

$$\mathrm{Q}_{aa} = \sum (\boldsymbol{a} - \bar{a})^2$$

$$= \sum \boldsymbol{a}^2 - \frac{\left(\sum \boldsymbol{a}\right)^2}{m} \tag{A.8}$$

$$\mathrm{Q}_{ca} = \sum (\boldsymbol{c} - \bar{c})\,(\boldsymbol{a} - \bar{a})$$

$$= \sum (\boldsymbol{ca}) - m\bar{c}\bar{a} \tag{A.9}$$

The correlation coefficient, r, is commonly used as a measure of the goodness of fit and can be calculated using equation (A.10). The R^2 statistic, which is commonly quoted with linear regression models, is the square of this value.

$$r = \frac{\mathrm{Q}_{ca}}{\sqrt{\mathrm{Q}_{cc}\,\mathrm{Q}_{aa}}} \tag{A.10}$$

Equation (A.11) is the expression for the calibration model when it is used for prediction of unknown ***c*** variables from measured ***a*** values.

$$\hat{\boldsymbol{c}} = \frac{\boldsymbol{a} - \hat{\mathrm{B}}_{0c}}{\hat{\mathrm{B}}_c} \tag{A.11}$$

A.2 Determination of univariate prediction intervals

Predictions intervals have been used since they provide an estimation of the uncertainty associated with the concentration of unknown samples. The confidence intervals characterise the uncertainty in determined ***a*** values.

The following treatment follows the International Union of Pure and Applied Chemistry (IUPAC) recommendations and is similar to that presented by Danzer and Currie.[1]

The residual standard deviation can be calculated using equation (A.12) where *m*-2 is equal to *f*, which is the degrees of freedom, in the case of a linear calibration model where the *y*-intercept is not equal to zero. When the *y*-intercept is zero then the degrees of freedom are increased to *m*-1.

$$\begin{aligned} s_{a,c} &= \sqrt{\frac{\sum(\boldsymbol{a} - \hat{\boldsymbol{a}})^2}{m\text{-}2}} \\ &= \sqrt{\frac{\sum(\boldsymbol{a} - \hat{\mathrm{B}}_0 - \hat{\mathrm{B}}\boldsymbol{c})^2}{m\text{-}2}} \end{aligned} \tag{A.12}$$

Once the residual standard deviation has been determined it can be used to calculate standard deviations for other parameters such as the slope and y-intercept (blank), shown in equations (A.13) and (A.14), respectively.

$$s_{\mathrm{B}_0} = s_{a,c}\sqrt{\frac{1}{m} + \frac{\bar{c}^2}{\mathrm{Q}_{cc}}} \tag{A.13}$$

$$s_B = \frac{s_{a,c}}{\sqrt{Q_{cc}}} \quad \text{(A.14)}$$

Standard deviations are not required for the calculation of the prediction intervals. Instead, the prediction interval for estimated $\boldsymbol{a_p}$ values using the linear calibration model over the concentration range $\boldsymbol{c_p}$ can be determined from equation (A.15), where $\boldsymbol{a_p}$ is determined using equation (A.3) with $\boldsymbol{c_p}$ being substituted for $\boldsymbol{c}$ and $t_{\alpha/2,f}$ is the Student's t statistic for f degrees of freedom (determined by the number of samples) and $1 - \alpha$ is the prediction level limit. Ninety-five percent prediction intervals are used throughout this work.

$$\text{prd}(a_p) = \boldsymbol{a_p} \pm s_{a,c}\, t_{\alpha/2,f} \sqrt{1 + \frac{1}{m} + \frac{(\boldsymbol{c_p} - \bar{c})^2}{Q_{cc}}} \quad \text{(A.15)}$$

A.3 References

1. Danzer, K. and Currie, L. A. Guidelines for calibration in analytical chemistry - Part 1. Fundamentals and single component calibration (IUPAC recommendations 1998), *Pure and Applied Chemistry*, **1998**, *70*, 993-1014.
2. Box, G. E. P., Hunter, W. G. and Hunter, J. S. *Statistics for Experimenters: An Introduction to Design Analysis and Model Building*; John Wiley & Sons: New York, 1978.

Appendix B

Quantification of P3 cosa PUR80

B.1 Quantification of P3 cosa® PUR80

The documentation provided by the manufacturer outlines two methods which may be suitable for the quantification of surface loadings prepared by the spray method.[1] The first is known as the "Curvette test" which is useful in the quantification of non-ionic polymeric surfactants. In this procedure a small amount ~2 mL of the aqueous washings which contains the P3 cosa® PUR80 cleaner is mixed with the same amount of an indicator, tetrabromophenolphthalein ethyl ester (TBPE) suspended in an organic solvent. The cleaner and the indicator form a complex which radiates at 605 nm, the intensity of this radiation is then monitored using a UV spectrometer and related to concentration in mg L^{-1}. The two major disadvantages of this technique are that it does require extraction of the complex into the organic phase and that interferences can be introduced by other surfactants and surface active ingredients present in the washing solutions.

The second technique is related to surface tension.[1] Again the concentration of P3 cosa® PUR80 is determined from aqueous washing solutions. Since the amount of P3 cosa® PUR80 affects the surface tension the concentration of the cleaner in a solution can be related to measurements of surface tension using a tensiometer. However this method is also subject to interferences which can be introduced by other surfactants and surface active ingredients present in the washing solutions, such as fatty substances which influence surface tension; it also requires the use of a tensiometer.

B.2 References

1. Ecolab, Cleaning Validation Handbook CD-Rom, Pharmocos, 2005